GPS-SUMO Manual

Prediction of SUMOylation Sites & SUMO-interaction Motifs

Version 1.0.1
05/10/2014

Author: Jian Ren & Yu Xue

Contact: Dr. Jian Ren, renjian.sysu@gmail.com; Dr. Yu Xue, xueyu@hust.edu.cn

The software is only free for academic research.
The latest version of GPS-SUMO software is available
from http://sumosp.biocuckoo.org

Index

Statement

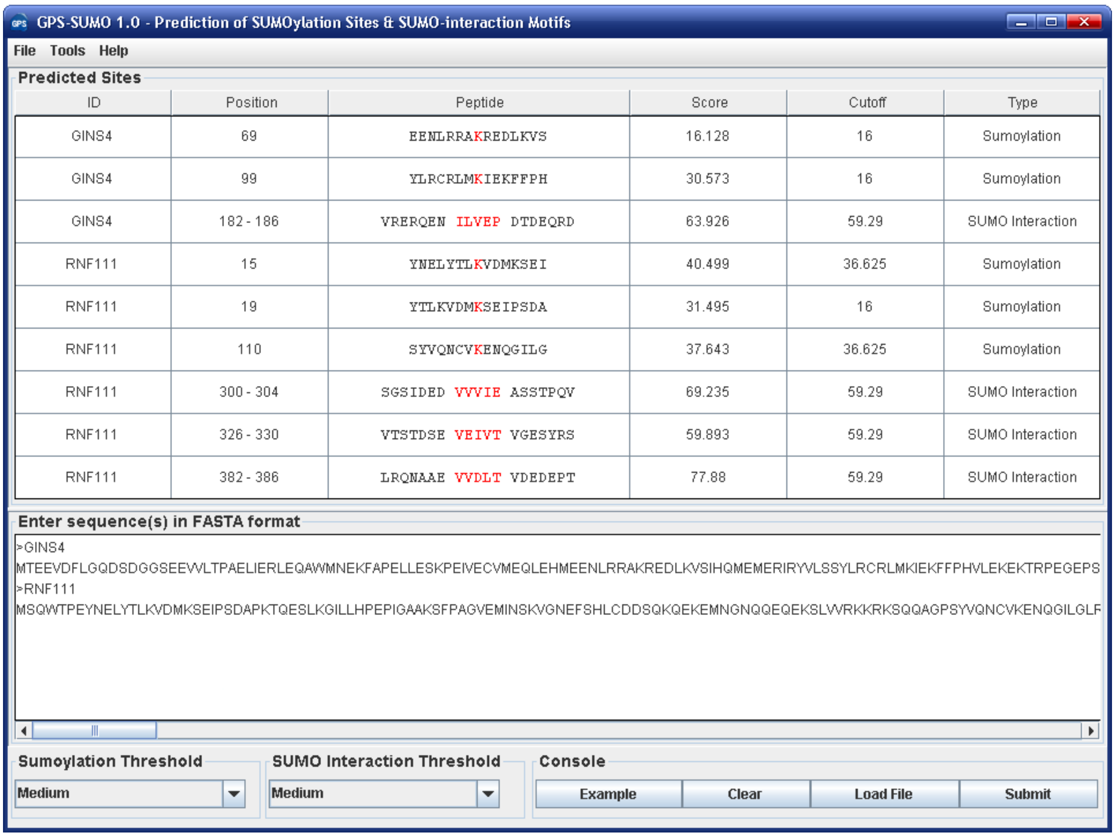
1. **Implementation**. The softwares of the CUCKOO Workgroup are implemented in JAVA (J2SE). Usually, both of online service and local stand-alone packages will be provided.
2. **Availability**. Our softwares are freely available for academic researches. For non-profit users, you can copy, distribute and use the softwares for your scientific studies. Our softwares are not free for commercial usage.
3. **GPS**. Previously, we used the GPS to denote our Group-based Phosphorylation Scoring algorithm. Currently, we are developing an integrated computational platform for post-translational modifications (PTMs) of proteins. We re-denote the GPS as Group-based Prediction Systems. This software is an indispensable part of GPS.
4. **Usage**. Our softwares are designed in an easy-to-use manner. Also, we invite you to read the manual before using the softwares.
5. **Updation**. Our softwares will be updated routinely based on users' suggestions and advices. Thus, your feedback is greatly important for our future updation. Please do not hesitate to contact with us if you have any concerns.
6. **Citation**. Usually, the latest published articles will be shown on the software websites. We wish you could cite the article if the software has been helpful for your work.
7. **Acknowledgements**. The work of CUCKOO Workgroup is supported by grants from the National Basic Research Program (973 project) [2013CB933902, 2012CB911201, 2012CB910101]; National Natural Science Foundation of China [31171263, 81272578]; the Guangdong Natural Science Funds for Distinguished Young Scholar [S20120011335]; Zhujiang Nova Program of Guangzhou [2011J2200042]; Program of International S&T Cooperation [2014DFB30020].

Introduction

Among the many protein post-translational modifications, sumoylation acts as a crucial biochemical process in the regulation of a variety of important biological functions. By specificly attaching a SUMO protein to a substrate, protein sumoylation could regulate multiple biochemical properties of protien target like the stability, activity, intracellular localization and protein interactions (1-3). Thousands of studies uncovered that sumoylation is essential for a serises of cellular processes, including DNA damage recovery, gene expression, chromosomal integrity as well as nuclear protein assemblies (4,5). In addition, protein sumoyaltion has shown to be intimately correlated with human diseases such as Alzheimer's disease (AD) (6), Parkinson's disease (PD)(7), viral infections(8), cardiac disease (9,10) and cancers(11).

Recently progresses revealed a class of SUMO-interaction motifs (SIMs) or SUMO-interacting motifs (SIMs), which mediate non-covalent interaction between SUMO and other proteins. The hydrophobic core of V/I-X-V/I-V/I or similar motifs provide additional specificity for sumoylation, and generate an interface for protein-protein interaction. However, such a motif will generate a huge number of potential hits for proteomic survey, which might be difficult for ambiguously experimental verifications. To data, there were only dozens of SIMs experimentally identified in proteins. In this regard, an accurate and efficient predictor is in urgent need for further experimental manipulation.

In this work, we reported an update of SUMOsp2.0 and renamed it to GPS-SUMO by mainly adding a novel SIM prediction feature. We first manually collected 151 known SIMs in 80 proteins from scientific literature. A new generation GPS (Group-based Prediction System) algorithm integrated with PSO (Particle Swarm Optimization) method was employed for predictor training. Due to the data limitation, only the leave-one-out validation was carried out to evaluate the prediction performance, with the sensitivity (Sn) of 92.7% and the specificity (Sp) of 95.0%. Additionally, to cover classic sumoylation sites prediction, the improved GPS algorithm and latest dataset were applied (986 sumoylation sites in 545 proteins). As the first computational tool for prediction of both SIMs and sumoylation sites, the web service of GPS-SUMO were implemented in JAVA and PHP, which is freely available at http://sumosp.biocuckoo.org/.


GPS-SUMO 1.0 - Prediction of SUMOylation Sites & SUMO-interaction Motifs
File
Tools
Help
Example
Clear
Load File
Submit

Predicted Sites

ID	Position	Peptide	Score	Cutoff	Type
GINS4	69	EENLRRAKREDLKVS	16.128	16	Sumoylation
GINS4	99	YLRCRLMKIEKFFPH	30.573	16	Sumoylation
GINS4	182 - 186	VRERQEN ILVEP DTDEQRD	63.926	59.29	SUMO Interaction
RNF111	15	YNELYTLKVDMKSEI	40.499	36.625	Sumoylation
RNF111	19	YTLKVDMKSEIPSDA	31.495	16	Sumoylation
RNF111	110	SYVQNCVKENQGILG	37.643	36.625	Sumoylation
RNF111	300 - 304	SGSIDED VVVIE ASSTPQV	69.235	59.29	SUMO Interaction
RNF111	326 - 330	VTSTDSE VEIVT VGESYRS	59.893	59.29	SUMO Interaction
RNF111	382 - 386	LRQNAAE VVDLT VDEDEPT	77.88	59.29	SUMO Interaction

Enter sequence(s) in FASTA format

```
>GINS4
MTEEVDFLGQDSDGGSEEVVLTPAELIERLEQAWMNEKFAPELLESKPEIVECVMEQLEHMEENLRRAKREDLKVSIHQMEMERIRYVLSSYLRCRLMKIEKFFPHVLEKEKTRPEGEPS
>RNF111
MSQWTPEYNELYTLKVDMKSEIPSDAPKTQESLKGILLHPEPIGAAKSFPAGVEMINSKVGNEFSHLCDDSQKQEKEMNGNQQEQEKSLVVRKKRKSQQAGPSYVQNCVKENQGILGLF
```

Sumoylation Threshold: Medium

SUMO Interaction Threshold: Medium

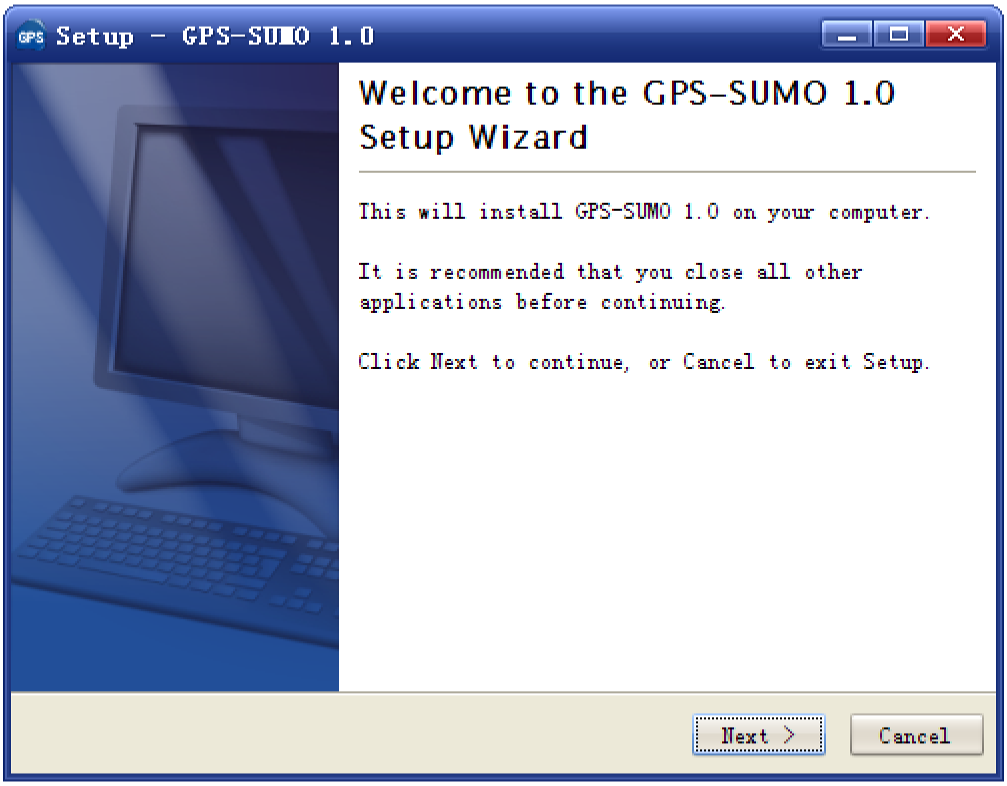
Console

GPS-SUMO 1.0 User Interface

Download & Installation

The GPS-SUMO was implemented in JAVA (J2SE), and could support three major Operating Systems (OS), including Windows, Linux/Unix or Mac OS X systems. Both of online web service and local stand-alone packages are available from: http://sumosp.biocuckoo.org/online.php. We recommend that users could download the latest release.

Please choose the proper package to download. After downloading, please double-click on the software package to begin installation, following the user prompts through the installation. And snapshots of the setup program for windows are shown below:

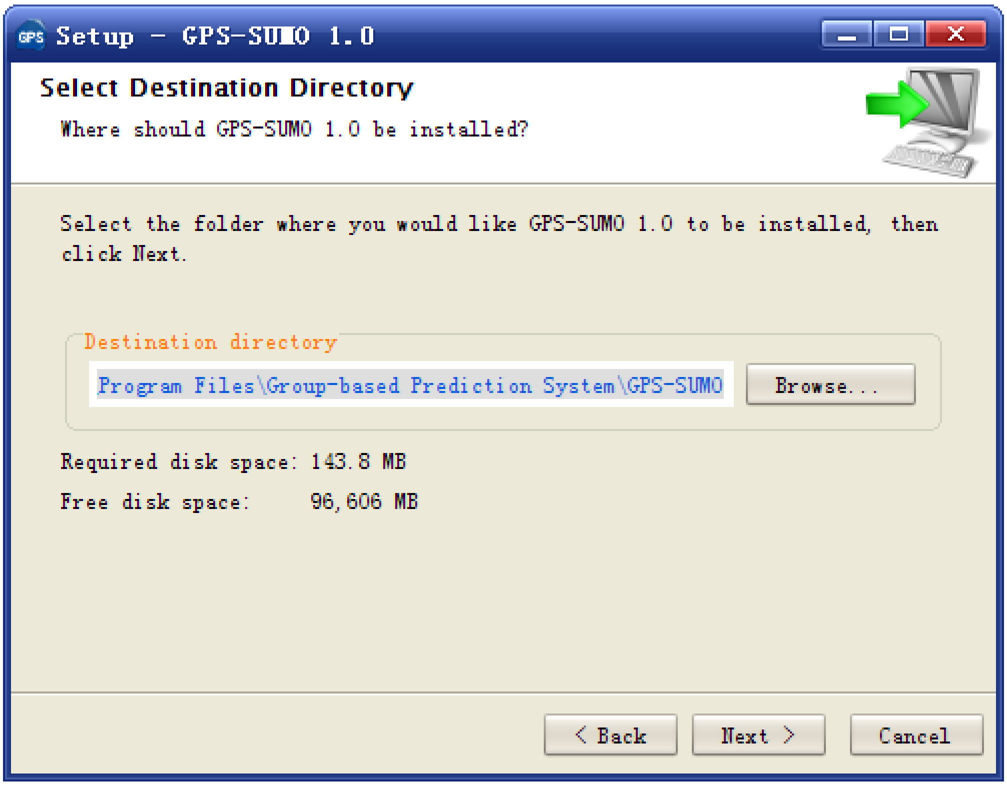


Welcome to the GPS-SUMO 1.0 Setup Wizard

This will install GPS-SUMO 1.0 on your computer.

It is recommended that you close all other applications before continuing.

Click Next to continue, or Cancel to exit Setup.



Setup - GPS-SUMO 1.0

Select Destination Directory

Where should GPS-SUMO 1.0 be installed?

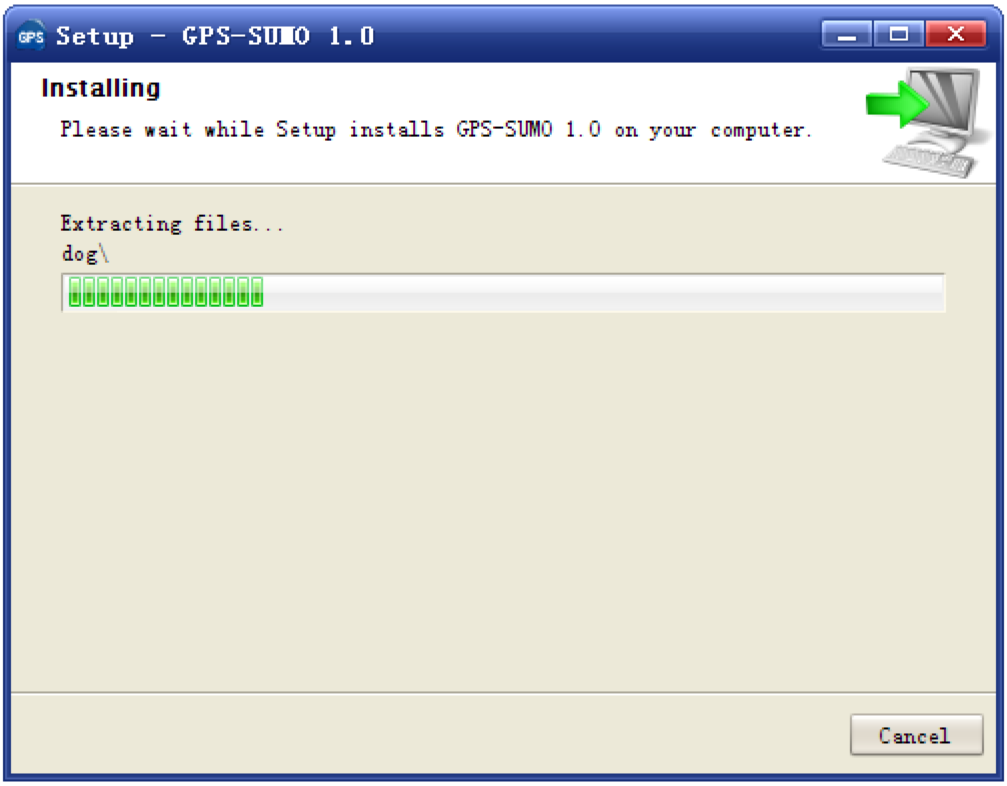
Select the folder where you would like GPS-SUMO 1.0 to be installed, then click Next.

Destination directory

Program Files\Group-based Prediction System\GPS-SUMO

Required disk space: 143.8 MB

Free disk space: 96,606 MB



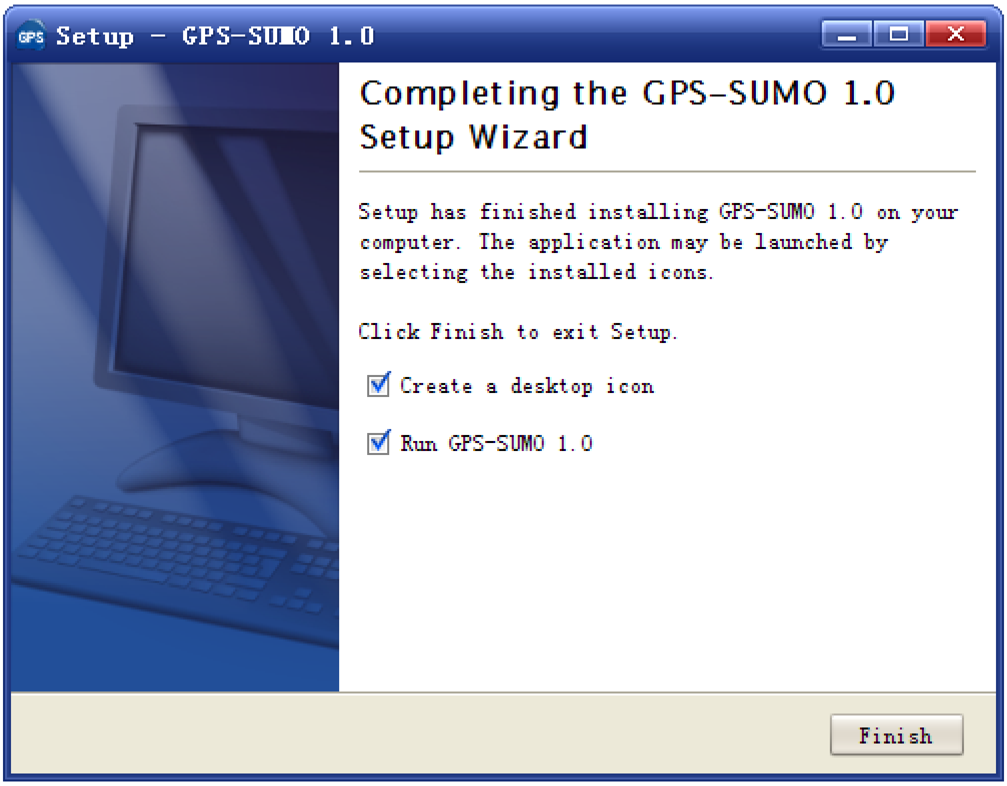
Setup - GPS-SUMO 1.0

Installing

Please wait while Setup installs GPS-SUMO 1.0 on your computer.

Extracting files...

dog\



Completing the GPS-SUMO 1.0 Setup Wizard

Setup has finished installing GPS-SUMO 1.0 on your computer. The application may be launched by selecting the installed icons.

Click Finish to exit Setup.

☑ Create a desktop icon

☑ Run GPS-SUMO 1.0

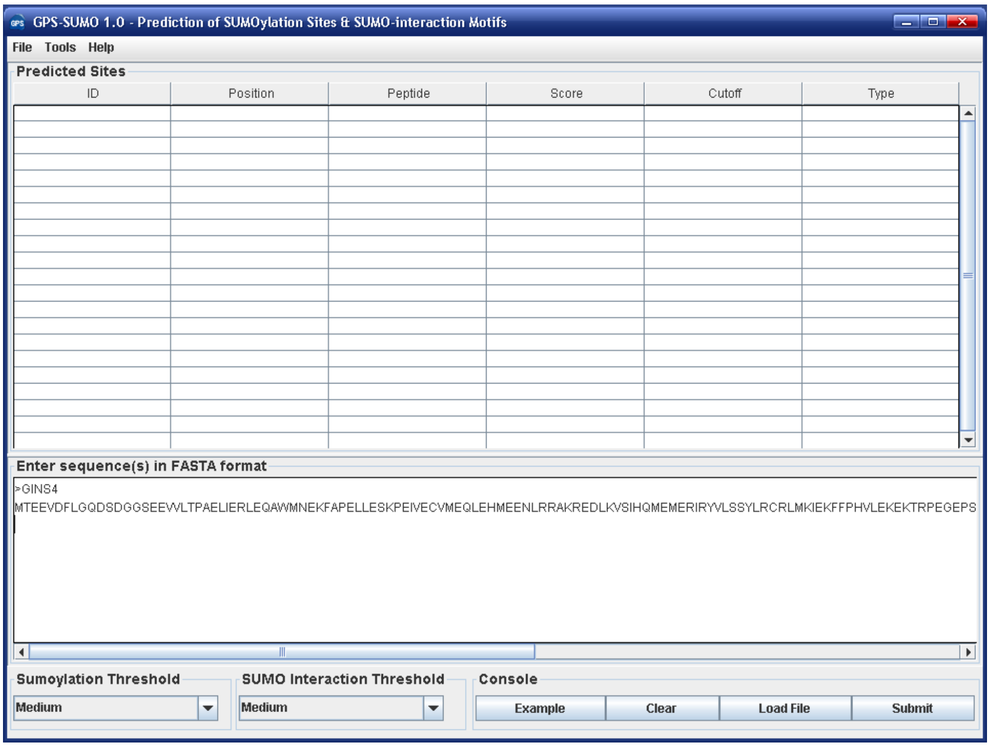
Finally, please click on the **Finish** button to complete the setup program.

Prediction of SUMO modification

A single protein sequence in FASTA format

The following steps show you how to use the GPS-SUMO 1.0 to predict SUMO modified sites for a single protein sequence in FASTA format.

(1) Firstly, please use “Ctrl+C & Ctrl+V” (Windows & Linux/Unix) or “Command+C & Command+V” (Mac) to copy and paste your sequence into the text form of GPS-SUMO 1.0


GPS-SUMO 1.0 - Prediction of SUMOylation Sites & SUMO-interaction Motifs
File
Tools
Help
Predicted Sites
ID
Position
Peptide
Score
Cutoff
Type
Enter sequence(s) in FASTA format
Sumoylation Threshold
SUMO Interaction Threshold
Console
Example
Clear
Load File
Submit

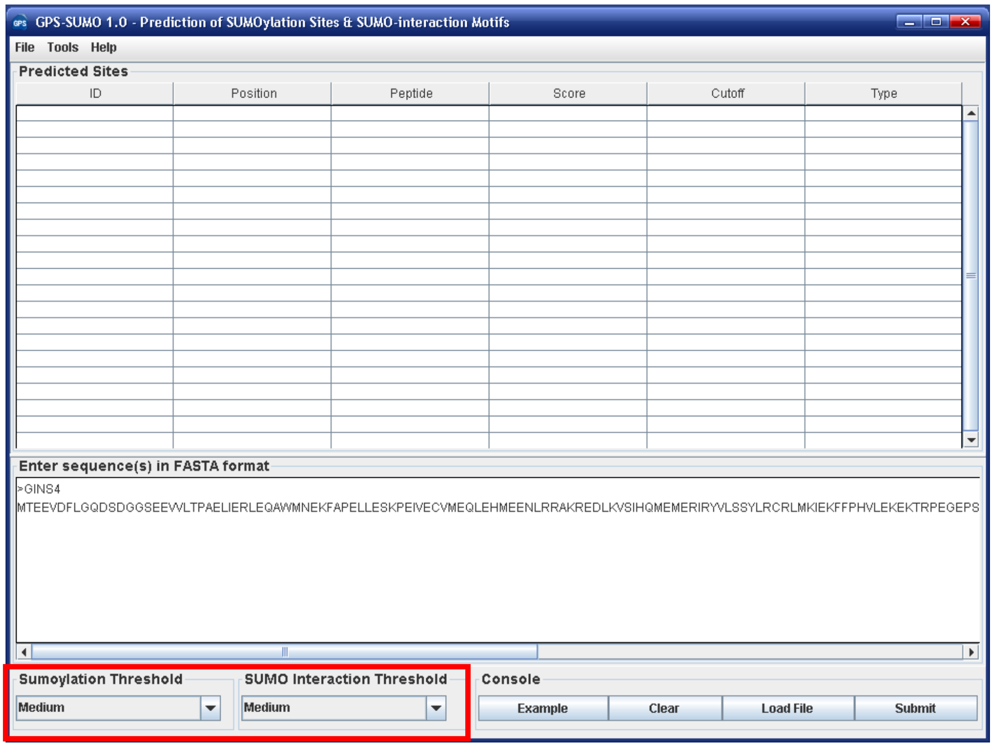
Sumoylation Threshold: Medium

SUMO Interaction Threshold: Medium

```
>GINS4
MTEEVDFLGQDSDGGSEEVVLTPAELIERLEQAWMNEKFAPELLESKPEIVECVMEQLEHMEENLRRAKREDLKVSIHQMEMERIRYVLSSYLRCRLMKIEKFFPHVLEKEKTRPEGEPS
```

Note: for a single protein, the sequence without a name in raw format is also OK. However, for multiple sequences, the name of each protein should be presented.

(2) Choose a **Thresholds** of Sumoylation and SUMO-interaction that you need, the default cut-off is **Medium**. Note that you can choose "None" to exclude the prediction of a corresponding modification type.



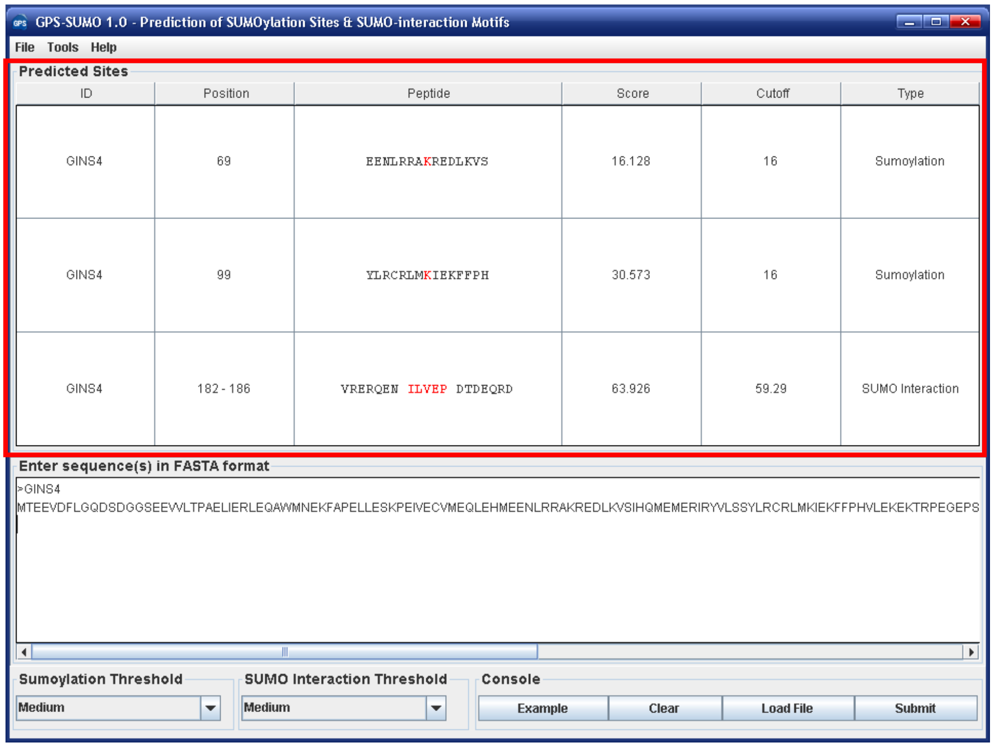
Enter sequence(s) in FASTA format

```
>GINS4
MTEEVDFLGQDSDGGSEEVVLTPAELIERLEQAWMNEKFAPELLESKPEIVECVMEQLEHMEENLRRAKREDLKVSIHQMEMERIRYVLSSYLRCRLMKIEKFFPHVLEKEKTRPEGEPS
```

Sumoylation Threshold: Medium

SUMO Interaction Threshold: Medium

(3) Click on the **Submit** button, then the predicted sumoylation sites/SUMO-interaction motif will be shown.


GPS-SUMO 1.0 - Prediction of SUMOylation Sites & SUMO-interaction Motifs
File Tools Help
Console
Example
Clear
Load File
Submit

Predicted Sites

ID	Position	Peptide	Score	Cutoff	Type
GINS4	69	EENLRRAKREDLKVS	16.128	16	Sumoylation
GINS4	99	YLRCRLMKIEKFFPH	30.573	16	Sumoylation
GINS4	182 - 186	VRERQEN ILVEP DTDEQRD	63.926	59.29	SUMO Interaction

Enter sequence(s) in FASTA format

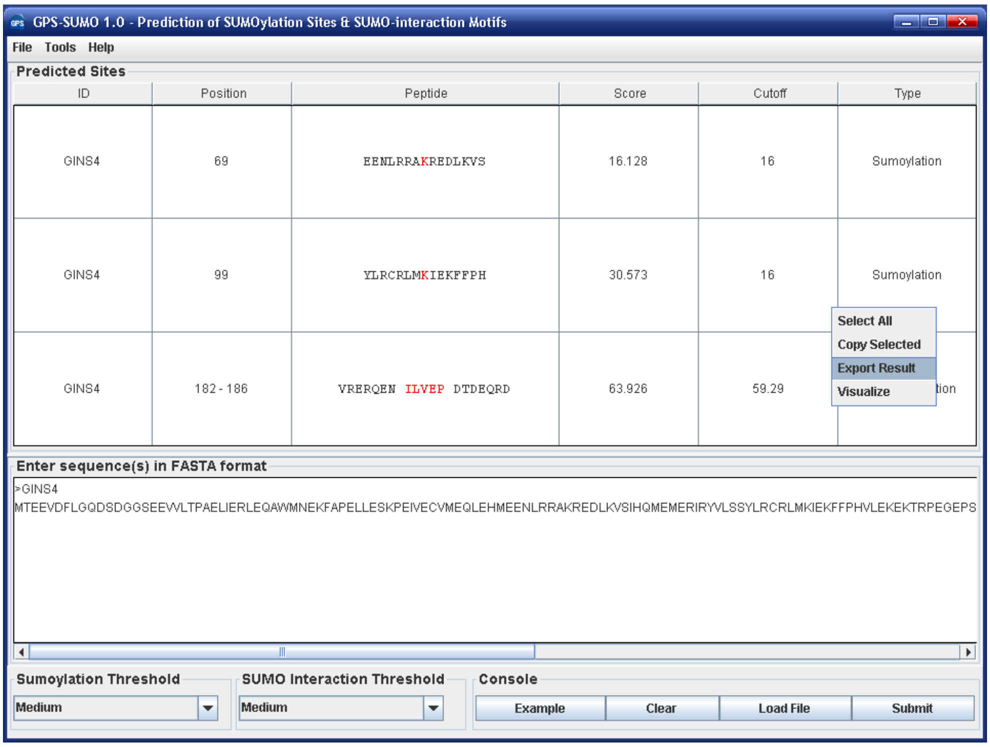
```
>GINS4
MTEEVDFLGQDSDGGSEEVVLTPAELIERLEQAWMNEKFAPELLESKPEIVECVMEQLEHMEENLRRAKREDLKVSIHQMEMERIRYVLSSYLRCRLMKIEKFFPHVLEKEKTRPEGEPS
```

Sumoylation Threshold: Medium

SUMO Interaction Threshold: Medium

(4) Then please click on the **RIGHT** button in the prediction form. You can use the

“**Select All**” and “**Copy Selected**” to copy the selected results into Clipboard. Then please copy the results into a file, eg., an EXCEL file for further consideration. Also, you can choose “**Export Prediction**” to export the prediction results into a tab-delimited text file.



Predicted Sites

ID	Position	Peptide	Score	Cutoff	Type
GINS4	69	EENLRRAKREDLKVS	16.128	16	Sumoylation
GINS4	99	YLRCRLMKIEKFFPH	30.573	16	Sumoylation
GINS4	182 - 186	VRERQEN ILVEP DTDEQRD	63.926	59.29	tion

Enter sequence(s) in FASTA format

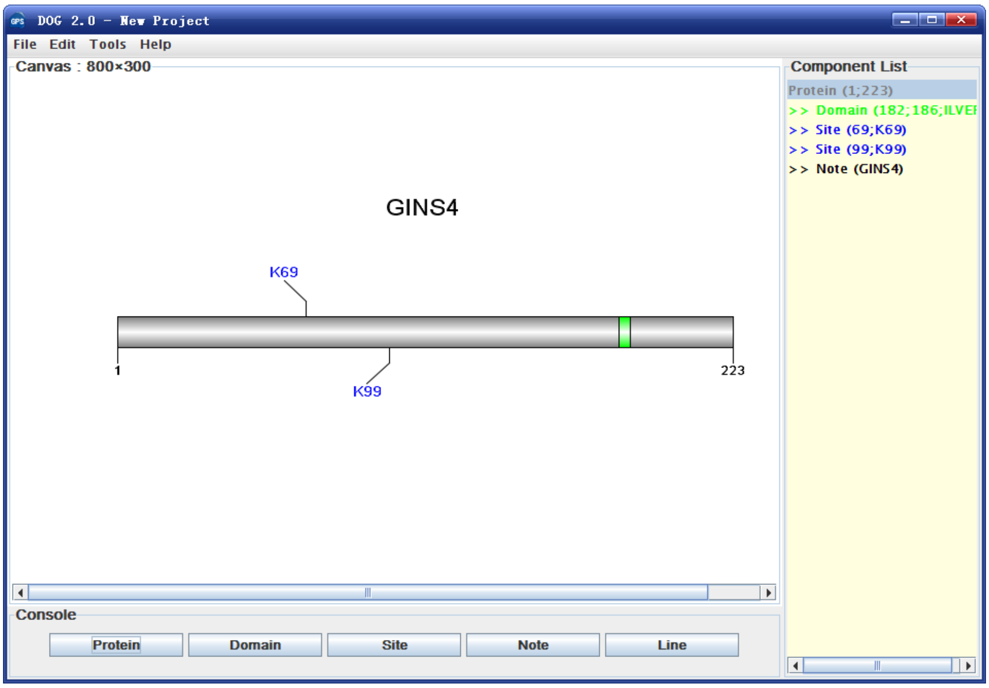
```
>GINS4
MTEEVDFLGQDSDGGSEEVVLTPAELIERLEQAWMNEKFAPELLESKPEIVECVMEQLEHMEENLRRAKREDLKVSIHQMEMERIRYVLSSYLRCRLMKIEKFFPHVLEKEKTRPEGEPS
```

Sumoylation Threshold: Medium

SUMO Interaction Threshold: Medium

Console

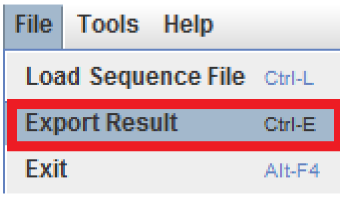
To visualize the predicted site please click on the “Visualize” menu.



Component List

Protein (1;223)
>> Domain (182;186;ILVEI
>> Site (69;K69)
>> Site (99;K99)
>> Note (GINS4)

Again, you can also click the “**Export Prediction**” in **File** menu to export the results.

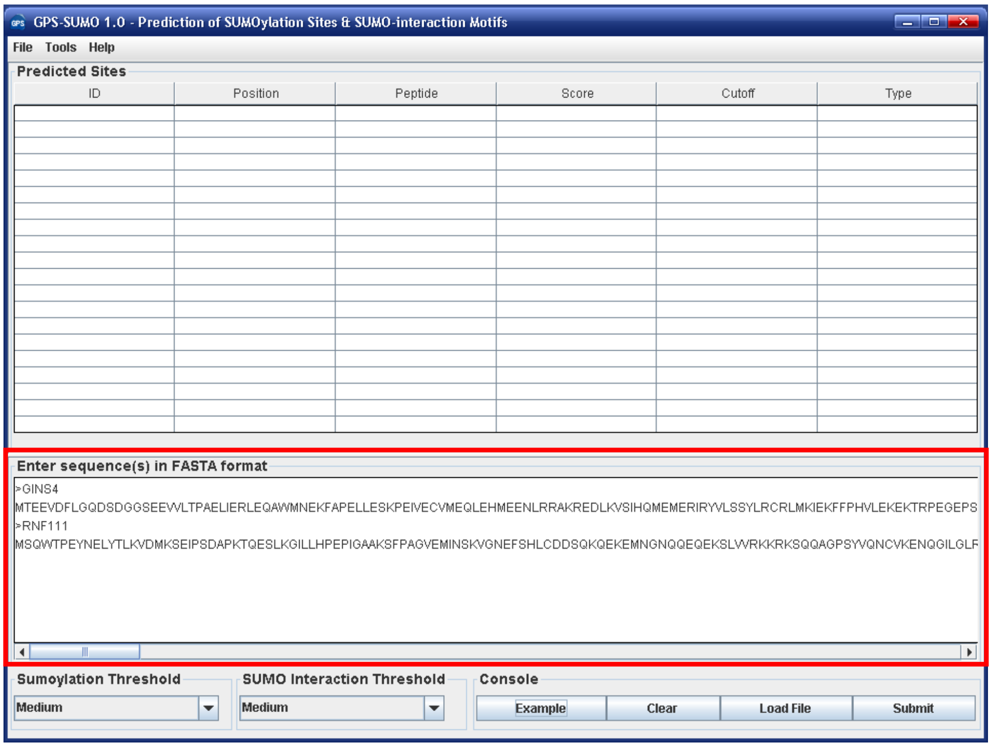


Multiple protein sequences in FASTA format

For multiple protein sequences, there are two ways to use the GPS-SUMO 1.0.

A. Input the sequences into text form directly. (Num. of Seq ≤2,000)

If the number of total protein sequences is not greater than 2,000, you can just use "Ctrl+C & Ctrl+V" (Windows & Linux/Unix) or "Command+C & Command+V" (Mac) to copy and paste your sequences into the text form of GPS-SUMO 1.0 for prediction.


GPS-SUMO 1.0 - Prediction of SUMOylation Sites & SUMO-interaction Motifs
File
Tools
Help
Predicted Sites
ID
Position
Peptide
Score
Cutoff
Type
Enter sequence(s) in FASTA format
Console
Example
Clear
Load File
Submit

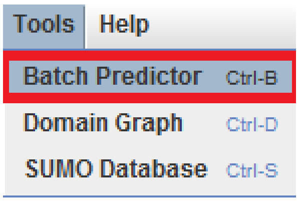
```
>GINS4
MTEEVDFLGQDSDGGSEEVVLTPAELIERLEQAWMNEKFAPELLESKPEIVECVMEQLEHMEENLRRAKREDLKVSIHQMEMERIRYVLSSYLRCRLMKIEKFFPHVLEKEKTRPEGEPS
>RNF111
MSQWTPEYNELYTLKVDMKSEIPSDAPKTQESLKGILLHPEPIGAAKSFPAGVEMINSKVGNEFSHLCDDSQKQEKEMNGNQQEQEKSLVVRKKRKSQQAGPSYVQNCVKENQGILGLF
```

Sumoylation Threshold: Medium

SUMO Interaction Threshold: Medium

B. Use Batch Predictor tool.

If the number of protein sequences is very large, eg., yeast or human proteome, please use the **Batch Predictor**. Please click on the "**Batch Predictor**" button in the **Tools** menu.



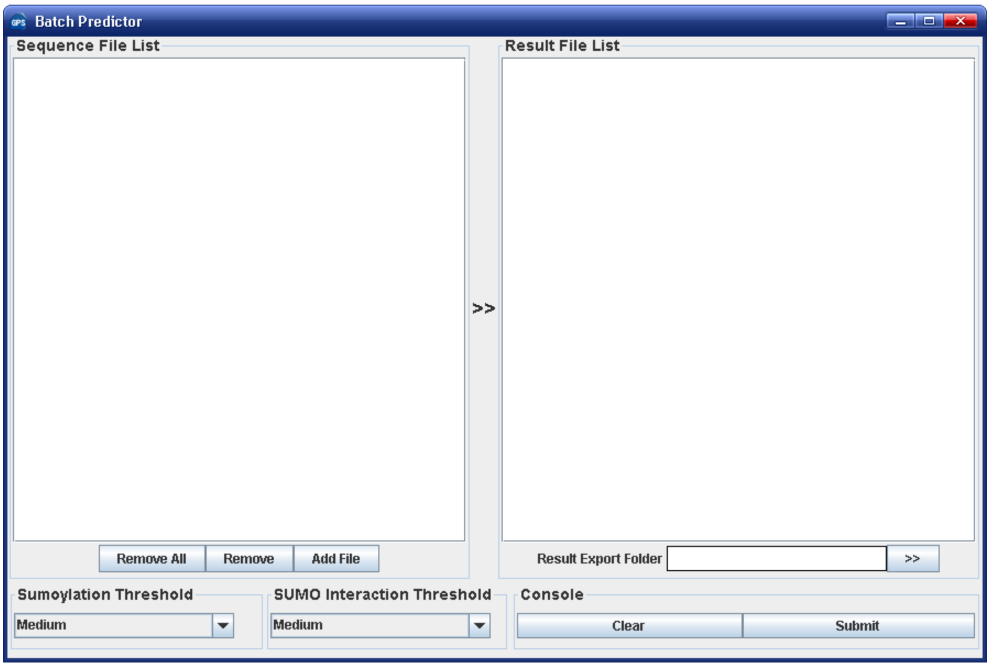
The following steps show you how to use it:

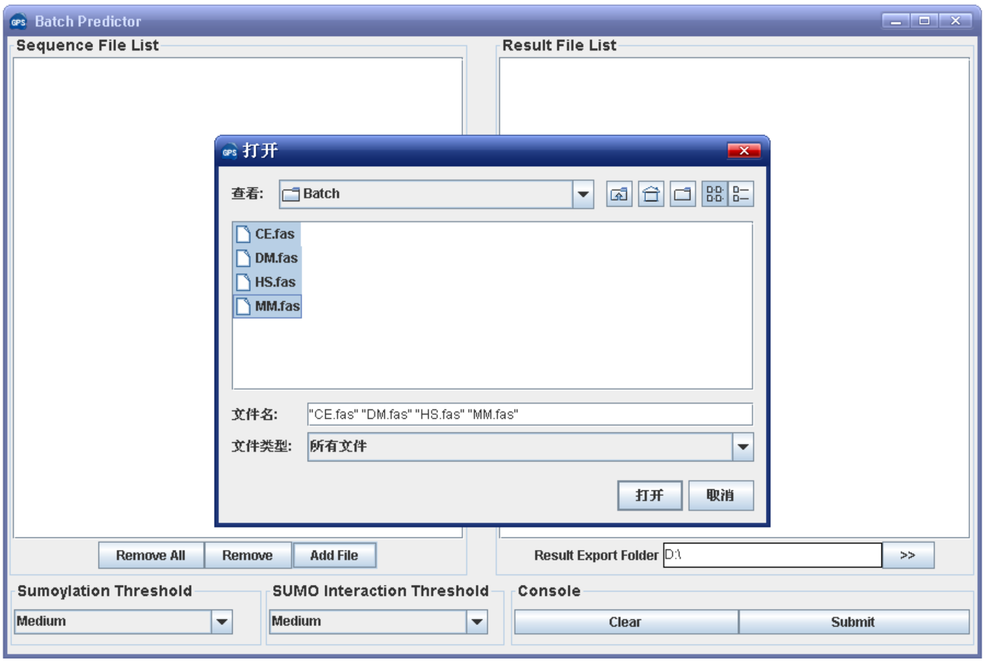
(1) Put protein sequences into one or several files (eg., SC.fas, CE.fas, and etc) with FATSA format as below:

```
>protein1
XXXXXXXXXXXXX
XXXXXXXX
>protein2
XXXXXXXXXXXXXXXX…
>protein3
XXXXXXXXXXXX
...
```

Most importantly, the name of each protein should be presented.

(2) Click on the **Batch Predictor** button and then click on the **Add File** button and add one or more protein sequence files in your hard disk.





查看: Batch

文件名: "CE.fas" "DM.fas" "HS.fas" "MM.fas"

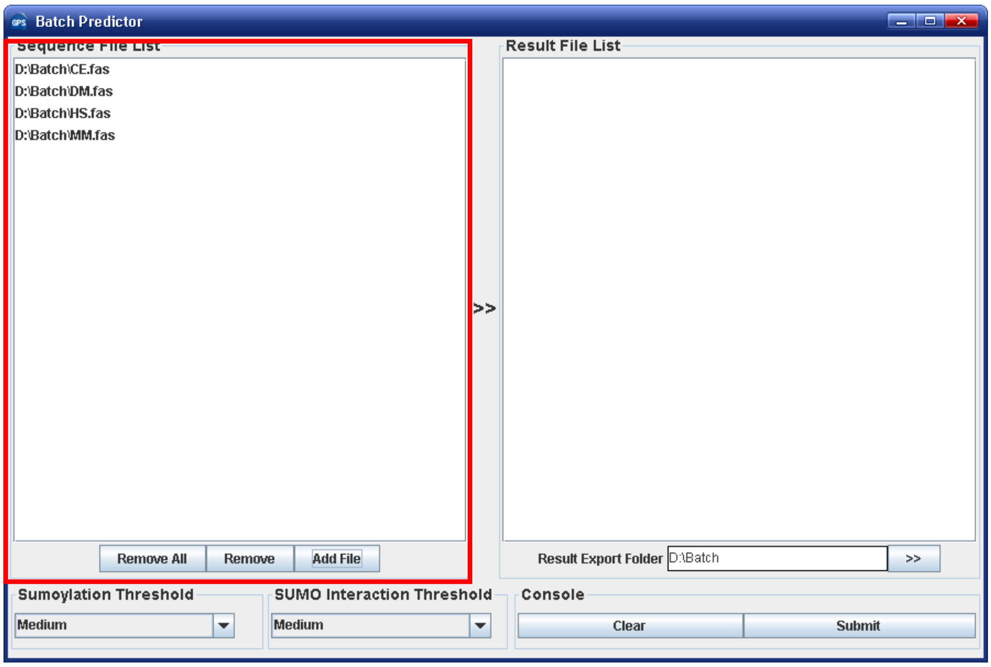
文件类型: 所有文件

Result Export Folder D:\

Sumoylation Threshold: Medium

SUMO Interaction Threshold: Medium

Then the names of added files will be shown in the **Sequence File List**.



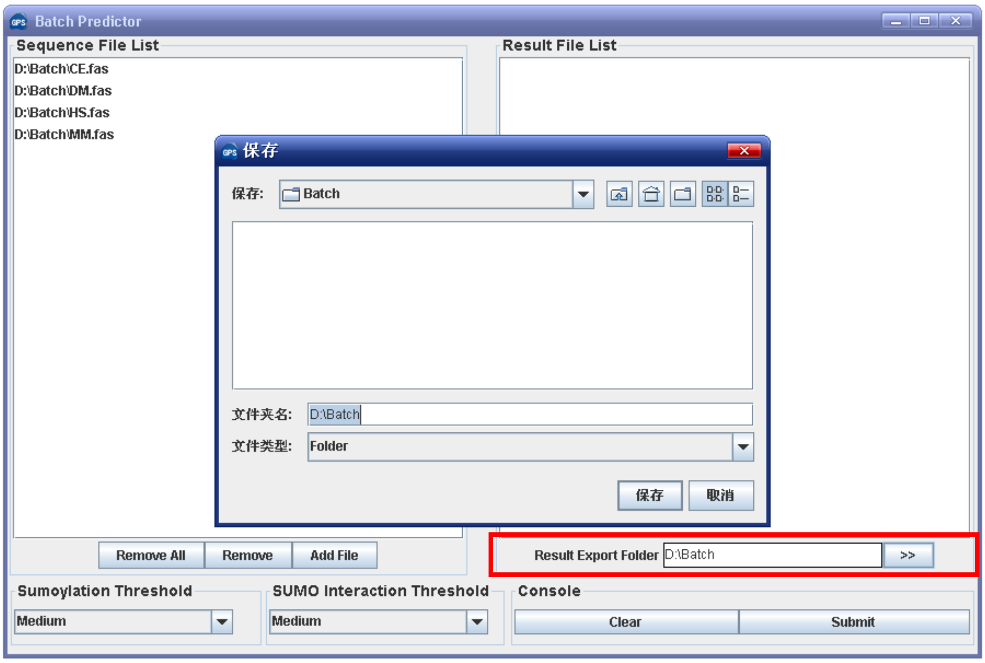
D:\Batch\CE.fas
D:\Batch\DM.fas
D:\Batch\HS.fas
D:\Batch\MM.fas

Result Export Folder D:\Batch

Sumoylation Threshold: Medium

SUMO Interaction Threshold: Medium

(3) The output directory of prediction results should also be defined. Please click on the >> button to specify the export fold.



Batch Predictor

Sequence File List

D:\Batch\CE.fas
D:\Batch\DM.fas
D:\Batch\HS.fas
D:\Batch\MM.fas

Result File List

保存

保存: Batch

文件夹名: D:\Batch

文件类型: Folder

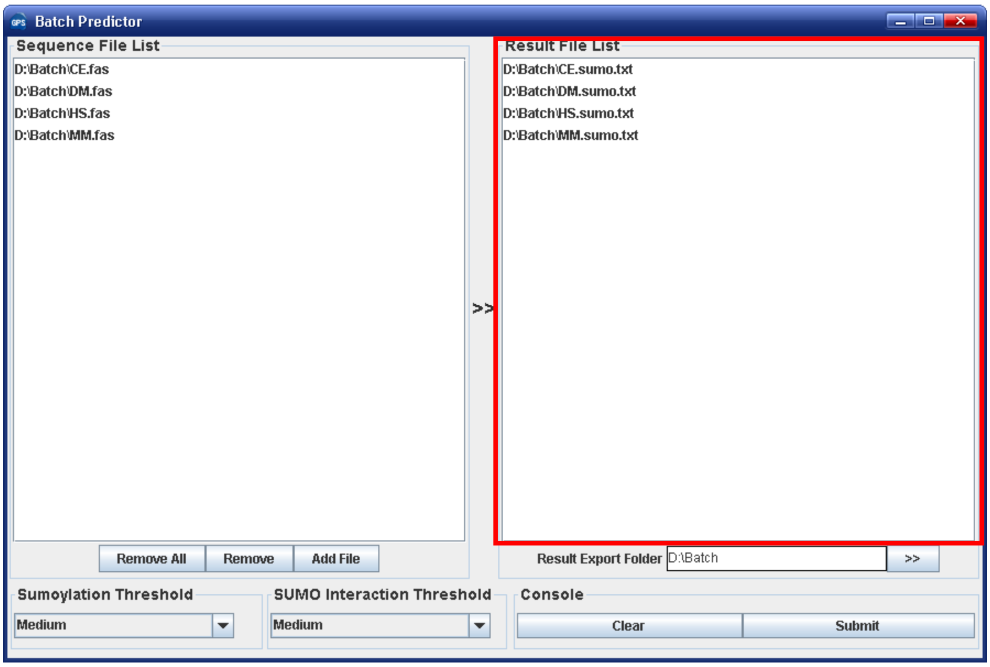
Result Export Folder D:\Batch

Sumoylation Threshold: Medium

SUMO Interaction Threshold: Medium

Console

(4) Please choose a proper threshold before prediction. Then please click on the **Submit** button, then the **Batch Predictor** begin to process all of the sequence files that have been added to the list. The result of prediction will be export to the **Result Export Fold,** and the name of result files will be shown in the **Result File List**.



Batch Predictor

Sequence File List

D:\Batch\CE.fas
D:\Batch\DM.fas
D:\Batch\HS.fas
D:\Batch\MM.fas

Result File List

D:\Batch\CE.sumo.txt
D:\Batch\DM.sumo.txt
D:\Batch\HS.sumo.txt
D:\Batch\MM.sumo.txt

Result Export Folder D:\Batch

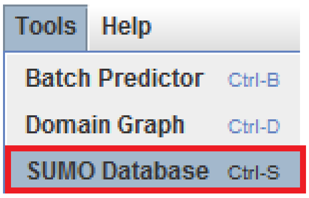
Sumoylation Threshold: Medium

SUMO Interaction Threshold: Medium

Console

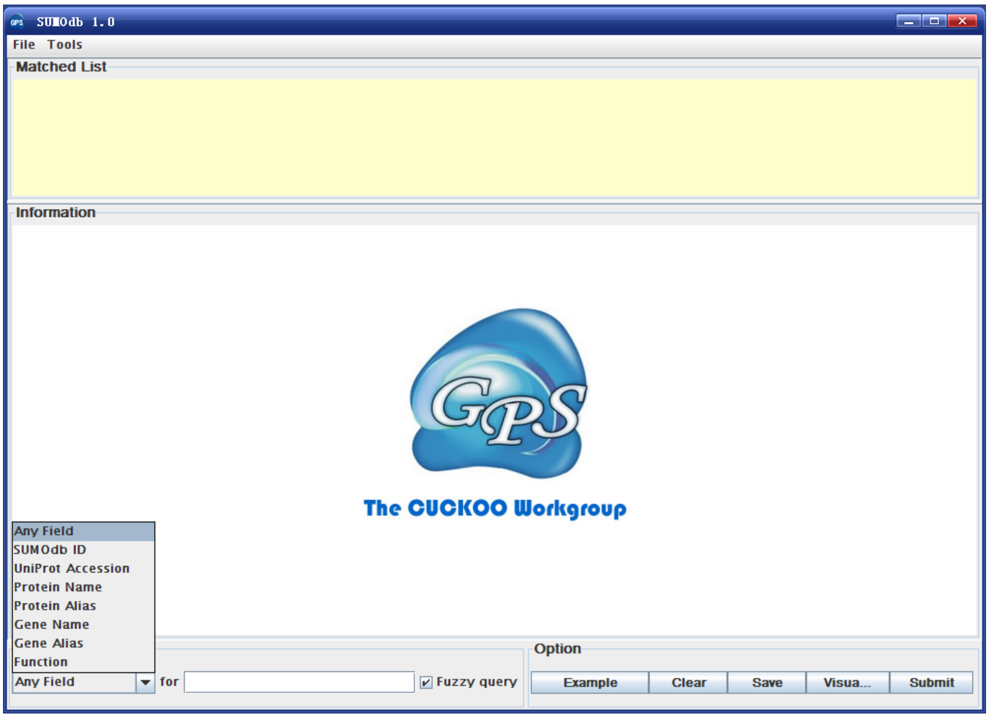
The usage of SUMO database

In GPS-SUMO 1.0, a SUMO modification database was integrated. Please click on the "**SUMO Database**" button in **Tool** menu to launch the database.

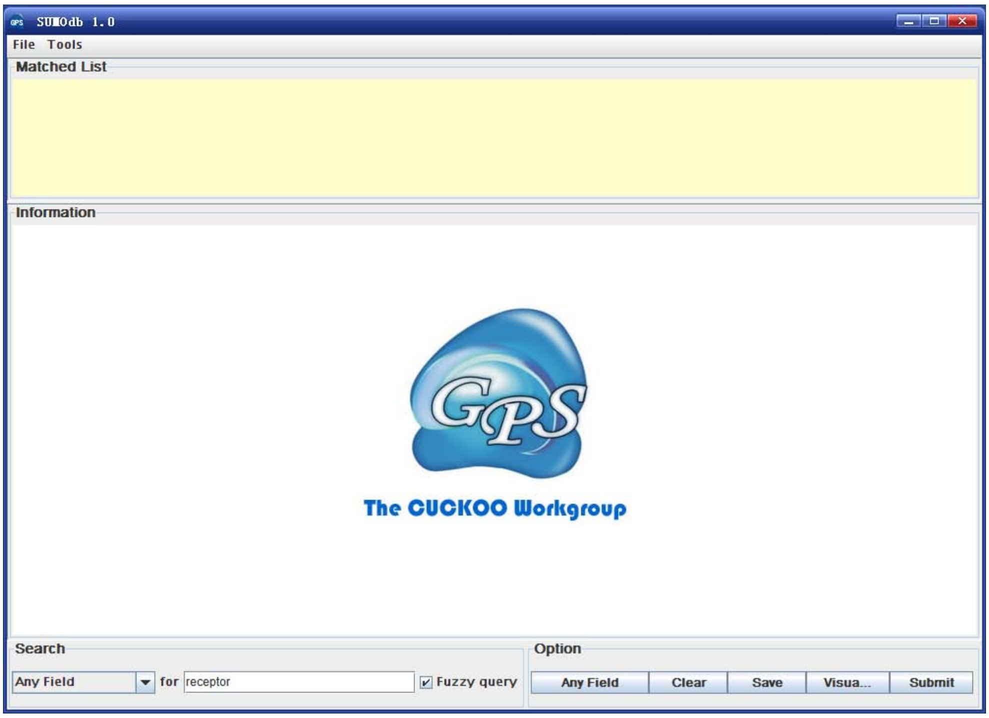


Search

The SUMO database was designed in an easy-to-use manner. For simple search, users could input a SUMO ID with SUMOdb-XXXX-XXXXXX, a UniProt ID (P02768), protein/gene names/aliases (eg., Serum albumin) or functions. Users could click the "Example" button one or several times to view the instances.

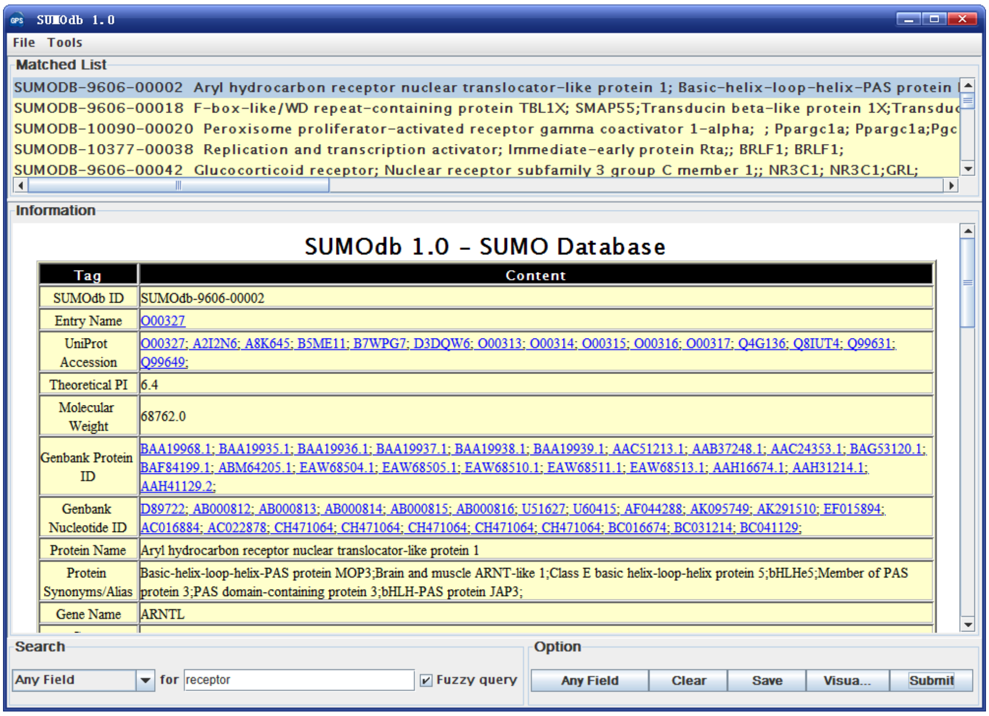


For example, users could input a protein/gene name/aliase, e.g., G-protein, specify the "**Any Field**", and then click on the "Submit" button to search the related information for this protein.



Any Field ▾ for receptor ☑ Fuzzy query

Then the information for G-protein will be shown in the “**Information**” form.



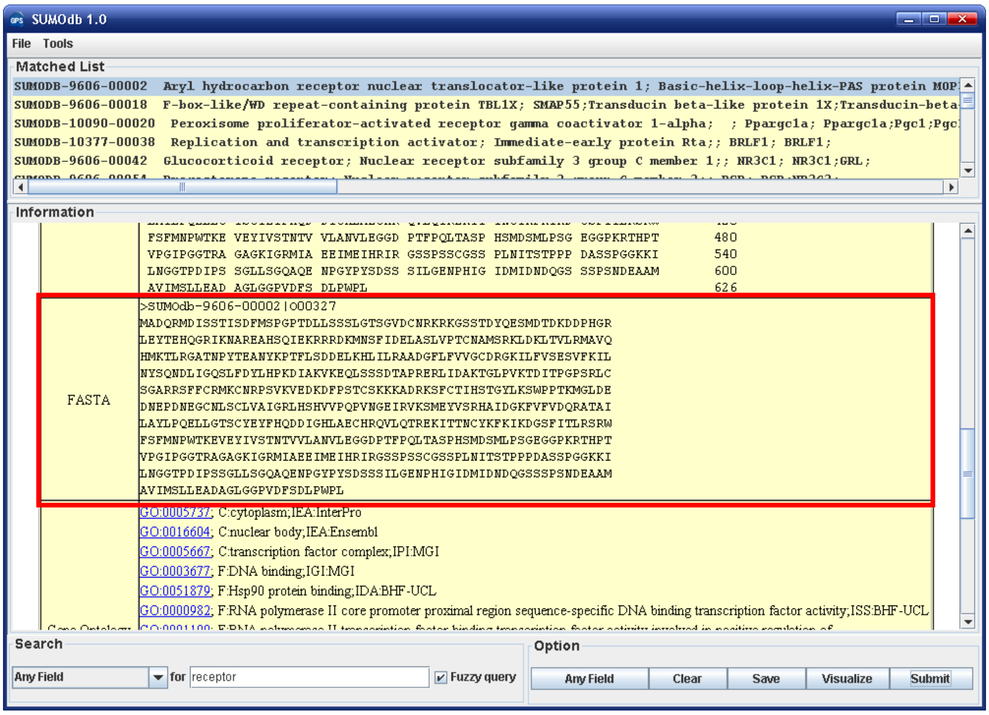
SUMODB-9606-00002 Aryl hydrocarbon receptor nuclear translocator-like protein 1; Basic-helix-loop-helix-PAS protein
SUMODB-9606-00018 F-box-like/WD repeat-containing protein TBL1X; SMAP55;Transducin beta-like protein 1X;Transduc
SUMODB-10090-00020 Peroxisome proliferator-activated receptor gamma coactivator 1-alpha; ; Ppargc1a; Ppargc1a;Pgc
SUMODB-10377-00038 Replication and transcription activator; Immediate-early protein Rta;; BRLF1; BRLF1;
SUMODB-9606-00042 Glucocorticoid receptor; Nuclear receptor subfamily 3 group C member 1;; NR3C1; NR3C1;GRL;

SUMOdb 1.0 – SUMO Database

Tag	Content
SUMOdb ID	SUMOdb-9606-00002
Entry Name	O00327
UniProt Accession	O00327; A2I2N6; A8K645; B5ME11; B7WPG7; D3DQW6; O00313; O00314; O00315; O00316; O00317; Q4G136; Q8IUT4; Q99631; Q99649;
Theoretical PI	6.4
Molecular Weight	68762.0
Genbank Protein ID	BAA19968.1; BAA19935.1; BAA19936.1; BAA19937.1; BAA19938.1; BAA19939.1; AAC51213.1; AAB37248.1; AAC24353.1; BAG53120.1; BAF84199.1; ABM64205.1; EAW68504.1; EAW68505.1; EAW68510.1; EAW68511.1; EAW68513.1; AAH16674.1; AAH31214.1; AAH41129.2;
Genbank Nucleotide ID	D89722; AB000812; AB000813; AB000814; AB000815; AB000816; U51627; U60415; AF044288; AK095749; AK291510; EF015894; AC016884; AC022878; CH471064; CH471064; CH471064; CH471064; CH471064; BC016674; BC031214; BC041129;
Protein Name	Aryl hydrocarbon receptor nuclear translocator-like protein 1
Protein Synonyms/Alias	Basic-helix-loop-helix-PAS protein MOP3;Brain and muscle ARNT-like 1;Class E basic helix-loop-helix protein 5;bHLHe5;Member of PAS protein 3;PAS domain-containing protein 3;bHLH-PAS protein JAP3;
Gene Name	ARNTL

Any Field ▾ for receptor ☑ Fuzzy query

The FASTA sequence for this protein would also be shown in the “**Information**” form.



```
SUMODB-9606-00002  Aryl hydrocarbon receptor nuclear translocator-like protein 1; Basic-helix-loop-helix-PAS protein MOP
SUMODB-9606-00018  F-box-like/WD repeat-containing protein TBL1X; SMAP55;Transducin beta-like protein 1X;Transducin-beta
SUMODB-10090-00020  Peroxisome proliferator-activated receptor gamma coactivator 1-alpha;  ; Ppargc1a; Ppargc1a;Pgc1;Pgc
SUMODB-10377-00038  Replication and transcription activator; Immediate-early protein Rta;; BRLF1; BRLF1;
SUMODB-9606-00042  Glucocorticoid receptor; Nuclear receptor subfamily 3 group C member 1;; NR3C1; NR3C1;GRL;
```

```
 FSFMNPWTKE VEYIVSTNTV VLANVLEGGD PTFPQLTASP HSMDSMLPSG EGGPKRTHPT       480
 VPGIPGGTRA GAGKIGRMIA EEIMEIHRIR GSSPSSCGSS PLNITSTPPP DASSPGGKKI       540
 LNGGTPDIPS SGLLSGQAQE NPGYPYSDSS SILGENPHIG IDMIDNDQGS SSPSNDEAAM       600
 AVIMSLLEAD AGLGGPVDFS DLPWPL                                            626
```

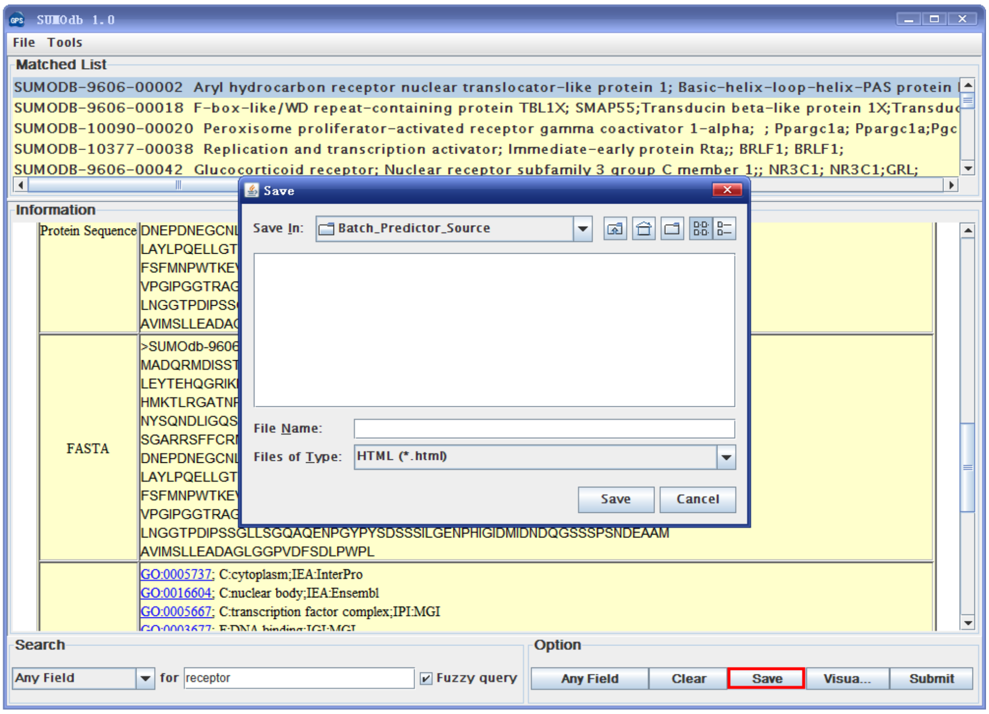
FASTA

```
>SUMOdb-9606-00002|O00327
MADQRMDISSTISDFMSPGPTDLLSSSLGTSGVDCNRKRKGSSTDYQESMDTDKDDPHGR
LEYTEHQGRIKNAREAHSQIEKRRRDKMNSFIDELASLVPTCNAMSRKLDKLTVLRMAVQ
HMKTLRGATNPYTEANYKPTFLSDDELKHLILRAADGFLFVVGCDRGKILFVSESVFKIL
NYSQNDLIGQSLFDYLHPKDIAKVKEQLSSSDTAPRERLIDAKTGLPVKTDITPGPSRLC
SGARRSFFCRMKCNRPSVKVEDKDFPSTCSKKKADRKSFCTIHSTGYLKSWPPTKMGLDE
DNEPDNEGCNLSCLVAIGRLHSHVVPQPVNGEIRVKSMEYVSRHAIDGKFVFVDQRATAI
LAYLPQELLGTSCYEYFHQDDIGHLAECHRQVLQTREKITTNCYKFKIKDGSFITLRSRW
FSFMNPWTKEVEYIVSTNTVVLANVLEGGDPTFPQLTASPHSMDSMLPSGEGGPKRTHPT
VPGIPGGTRAGAGKIGRMIAEEIMEIHRIRGSSPSSCGSSPLNITSTPPPDASSPGGKKI
LNGGTPDIPSSGLLSGQAQENPGYPYSDSSSILGENPHIGIDMIDNDQGSSSPSNDEAAM
AVIMSLLEADAGLGGPVDFSDLPWPL
```

GO:0005737; C:cytoplasm;IEA:InterPro
GO:0016604; C:nuclear body;IEA:Ensembl
GO:0005667; C:transcription factor complex;IPI:MGI
GO:0003677; F:DNA binding;IGI:MGI
GO:0051879; F:Hsp90 protein binding;IDA:BHF-UCL
GO:0000982; F:RNA polymerase II core promoter proximal region sequence-specific DNA binding transcription factor activity;ISS:BHF-UCL

Any Field ▾ for receptor ☑ Fuzzy query

The searched results could be saved in HTML format.



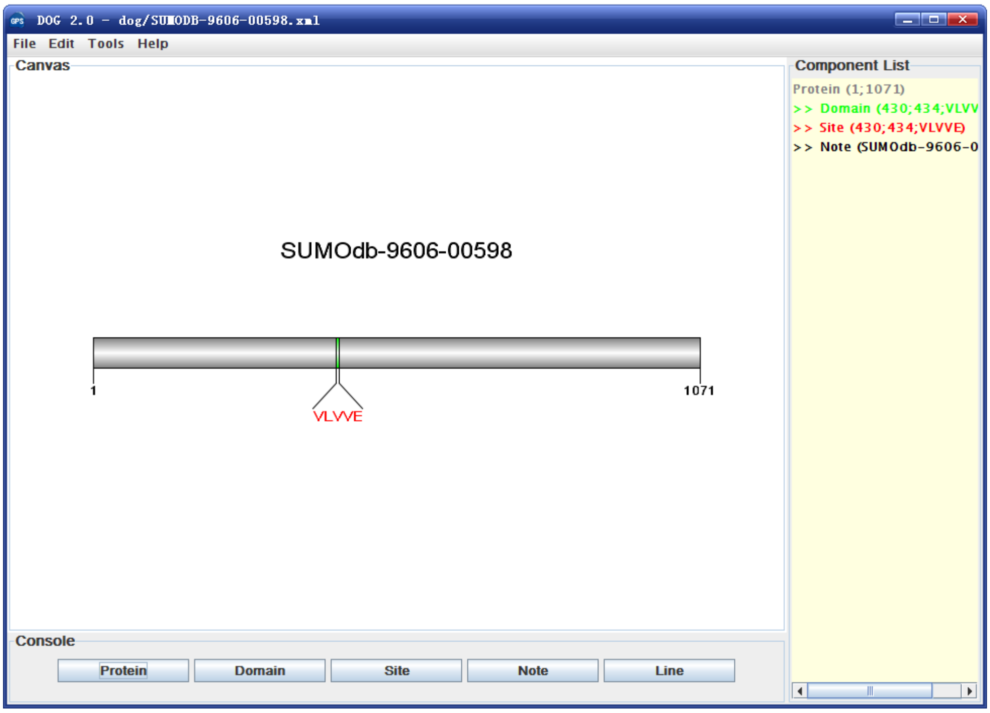
```
SUMODB-9606-00002  Aryl hydrocarbon receptor nuclear translocator-like protein 1; Basic-helix-loop-helix-PAS protein
SUMODB-9606-00018  F-box-like/WD repeat-containing protein TBL1X; SMAP55;Transducin beta-like protein 1X;Transduc
SUMODB-10090-00020  Peroxisome proliferator-activated receptor gamma coactivator 1-alpha;  ; Ppargc1a; Ppargc1a;Pgc
SUMODB-10377-00038  Replication and transcription activator; Immediate-early protein Rta;; BRLF1; BRLF1;
SUMODB-9606-00042  Glucocorticoid receptor; Nuclear receptor subfamily 3 group C member 1;; NR3C1; NR3C1;GRL;
```

```
LNGGTPDIPSSGLLSGQAQENPGYPYSDSSSILGENPHIGIDMIDNDQGSSSPSNDEAAM
AVIMSLLEADAGLGGPVDFSDLPWPL
```

GO:0005737; C:cytoplasm;IEA:InterPro
GO:0016604; C:nuclear body;IEA:Ensembl
GO:0005667; C:transcription factor complex;IPI:MGI

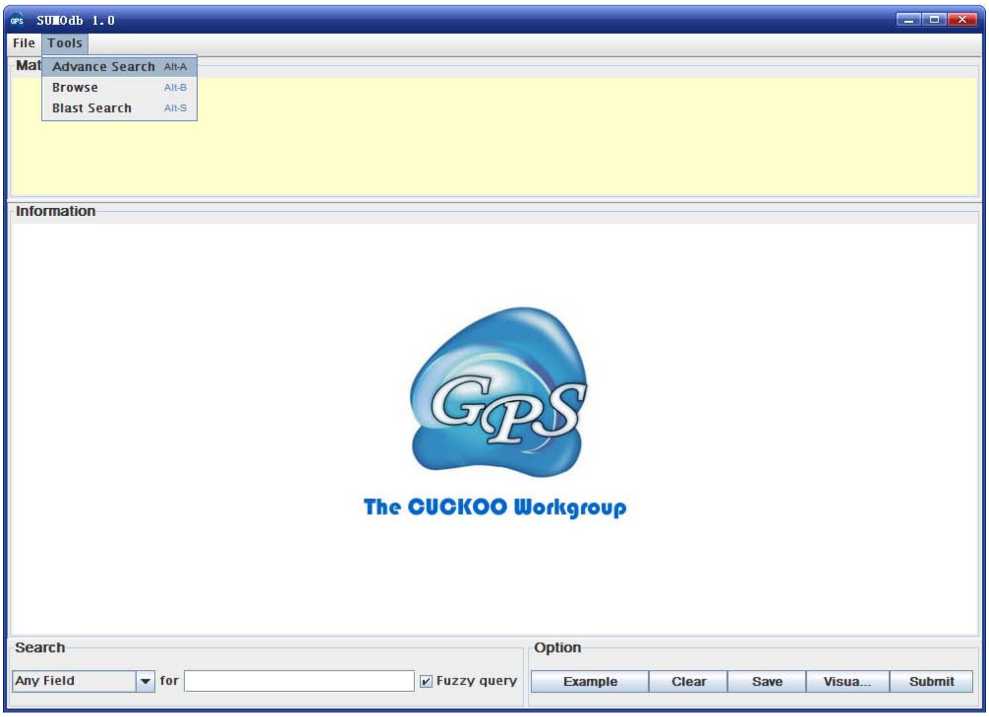
Any Field ▾ for receptor ☑ Fuzzy query

Notably, you can click on the “**Visualize**” button to view the sumoyaltion site or SUMO-interaction motif (green in graph) in our DOG 2 package.



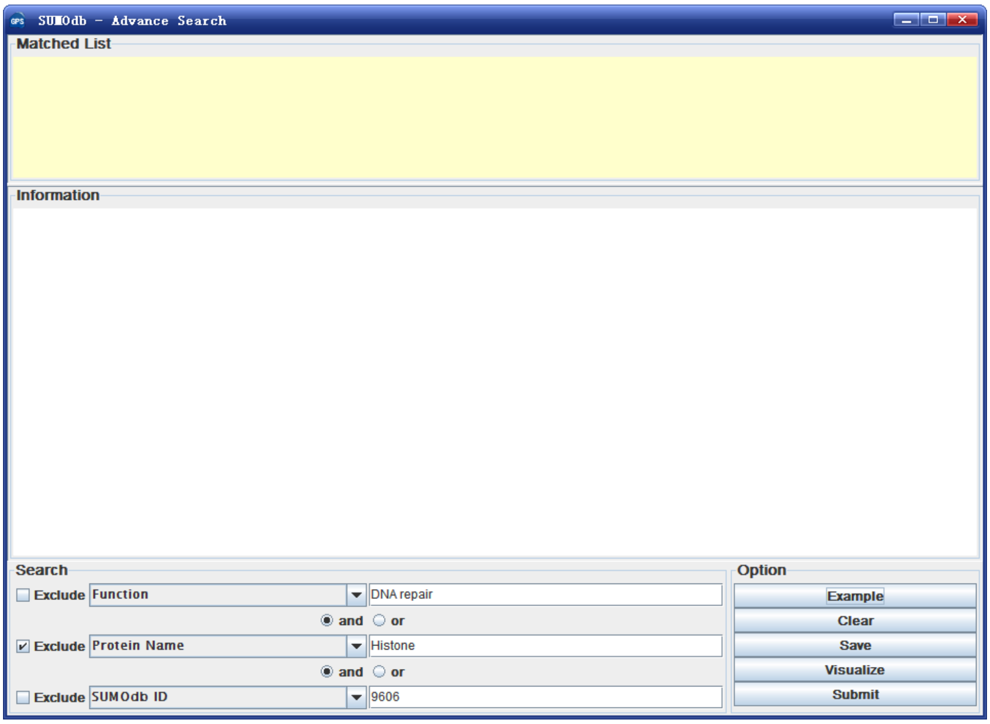
Advance search

The SUMO database supports three advance options, including advance search, browse, BLAST search. The Advance search option allows you to input up to three terms to find the information more specifically. The querying fields can be empty if fewer terms are needed.



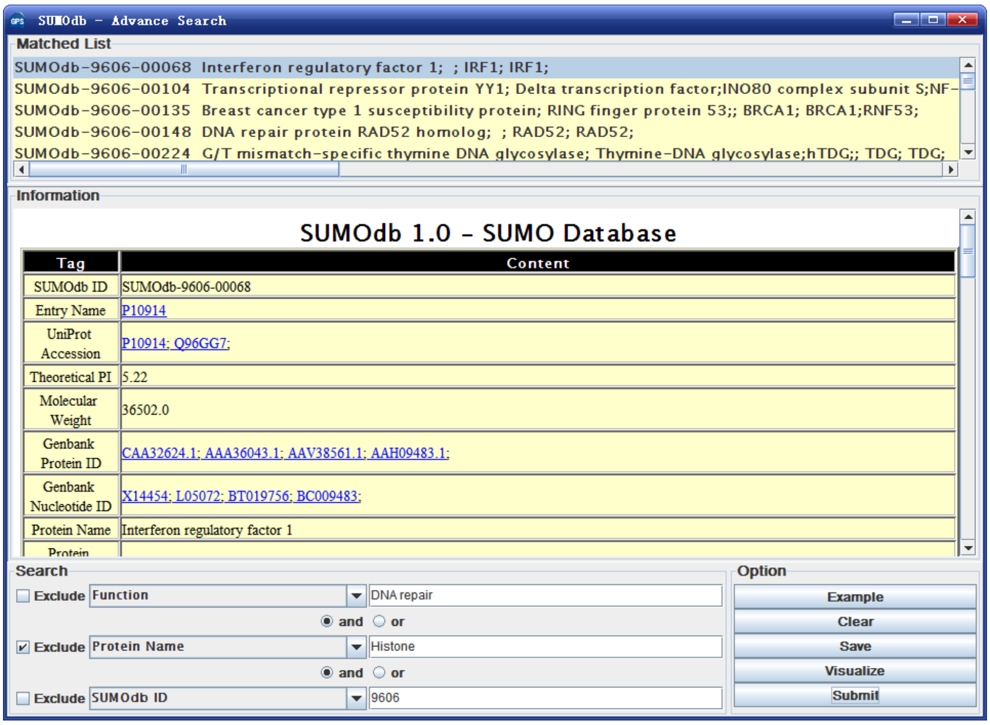
First, users could click on the “**Tools**” button then click on the “**Advance Search**” button to open this option.

By clicking the “Example” button, you can try an instance for usage. You can input DNA repair(Function), Histone (Protein Name), and 9606 (SUMOdb ID) for querying.



□ Exclude Function: DNA repair
● and ○ or
☑ Exclude Protein Name: Histone
● and ○ or
□ Exclude SUMOdb ID: 9606

Then the result will be shown as follow:



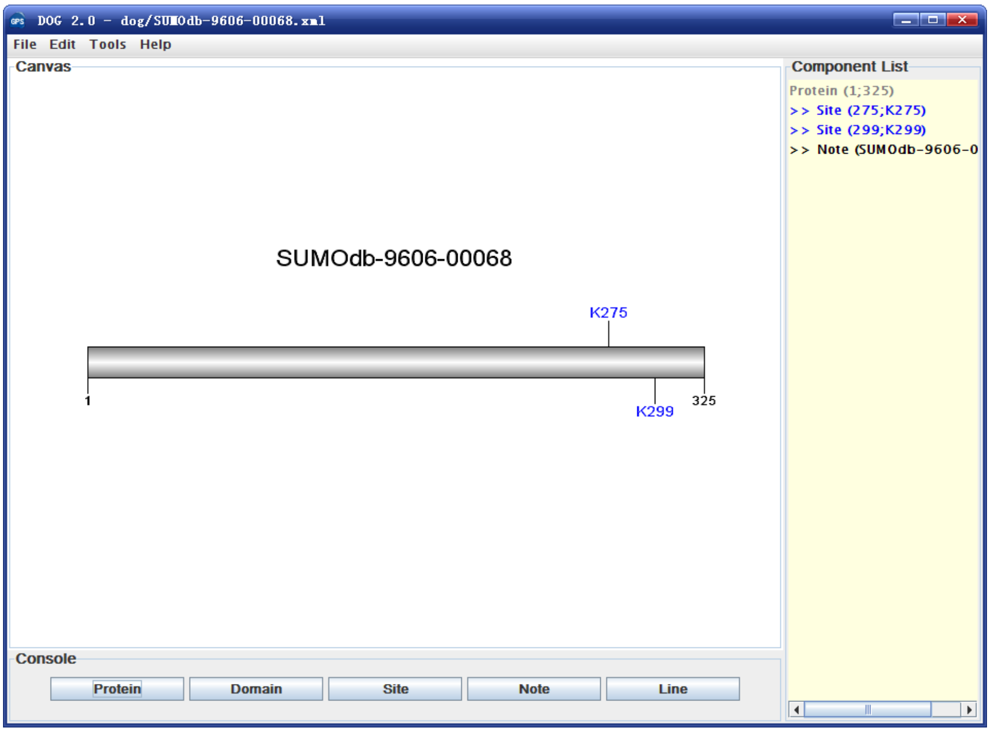
```
SUMOdb-9606-00068  Interferon regulatory factor 1;  ; IRF1; IRF1;
SUMOdb-9606-00104  Transcriptional repressor protein YY1; Delta transcription factor;INO80 complex subunit S;NF-
SUMOdb-9606-00135  Breast cancer type 1 susceptibility protein; RING finger protein 53;; BRCA1; BRCA1;RNF53;
SUMOdb-9606-00148  DNA repair protein RAD52 homolog;  ; RAD52; RAD52;
SUMOdb-9606-00224  G/T mismatch-specific thymine DNA glycosylase; Thymine-DNA glycosylase;hTDG;; TDG; TDG;
```

SUMOdb 1.0 - SUMO Database

Tag	Content
SUMOdb ID	SUMOdb-9606-00068
Entry Name	P10914
UniProt Accession	P10914; Q96GG7;
Theoretical PI	5.22
Molecular Weight	36502.0
Genbank Protein ID	CAA32624.1; AAA36043.1; AAV38561.1; AAH09483.1;
Genbank Nucleotide ID	X14454; L05072; BT019756; BC009483;
Protein Name	Interferon regulatory factor 1
Protein	

□ Exclude Function: DNA repair
● and ○ or
☑ Exclude Protein Name: Histone
● and ○ or
□ Exclude SUMOdb ID: 9606

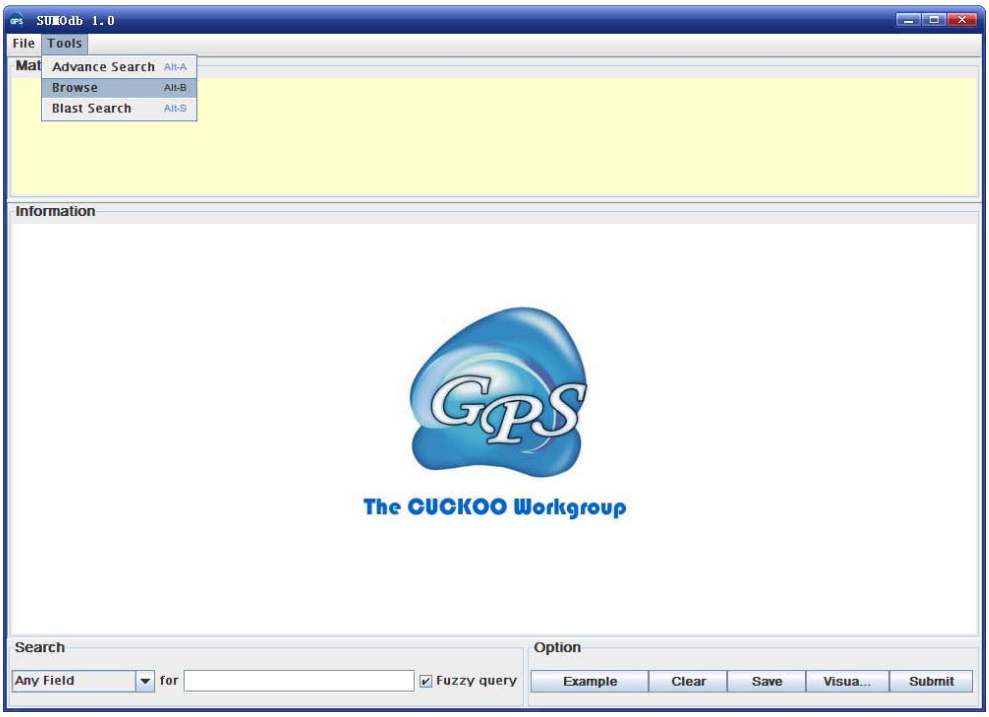
Again, you can click on the "**Visualize**" button to open a schematic diagram for sumoylation site.

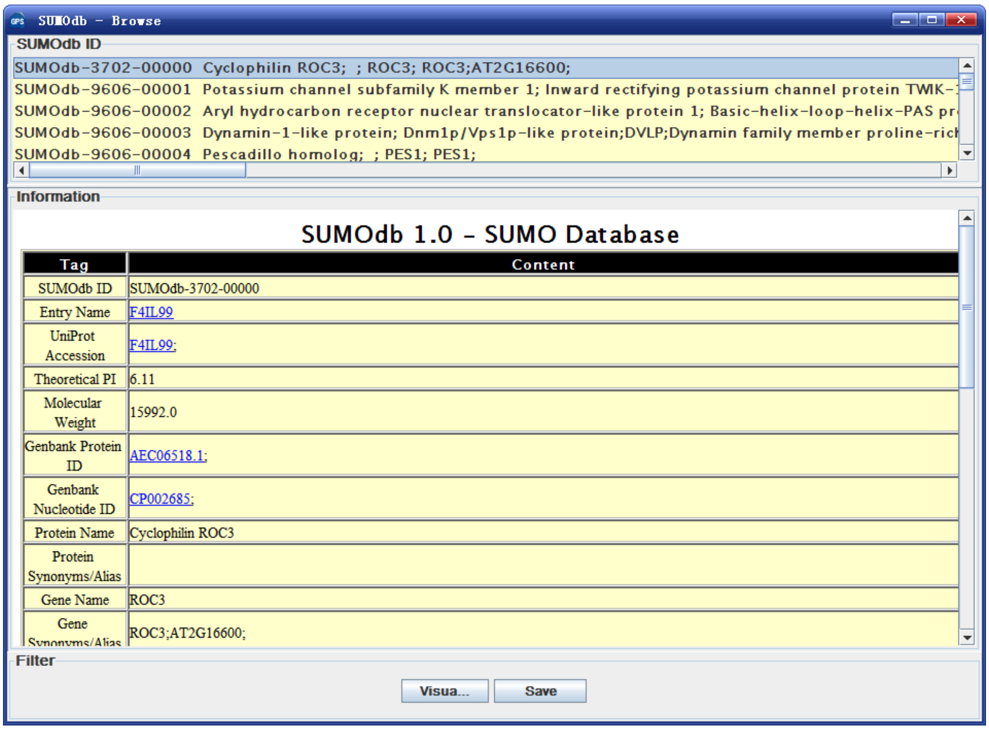


Browse

The SUMO database supports the browse function. The Browse search allows users to view all entries in SUMO database.

First, users could click on the "Tools" button then click on the "**Browse**" button to visualize all SUMO proteins. Users could visualize any protein by click on the entries Listed in the "SUMO ID" form. Also, by clicking the "Visualize" button, a schematic diagram of protein SUMO modified sites will be shown.





```
SUMOdb-3702-00000  Cyclophilin ROC3;  ; ROC3; ROC3;AT2G16600;
SUMOdb-9606-00001  Potassium channel subfamily K member 1; Inward rectifying potassium channel protein TWIK-
SUMOdb-9606-00002  Aryl hydrocarbon receptor nuclear translocator-like protein 1; Basic-helix-loop-helix-PAS pr
SUMOdb-9606-00003  Dynamin-1-like protein; Dnm1p/Vps1p-like protein;DVLP;Dynamin family member proline-rich
SUMOdb-9606-00004  Pescadillo homolog;  ; PES1; PES1;
```

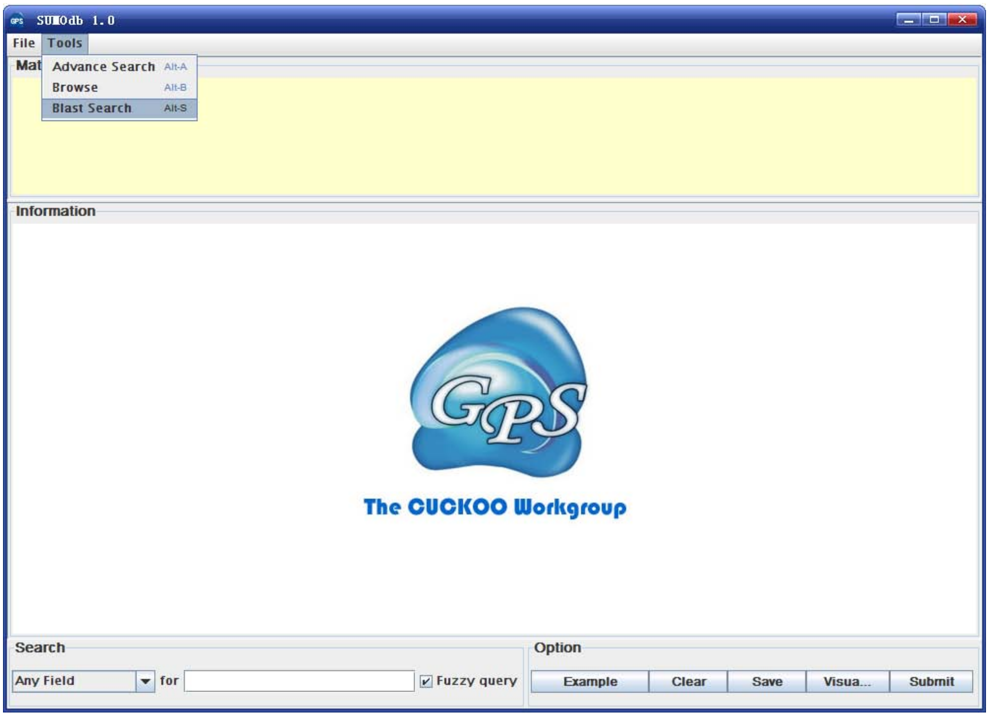
SUMOdb 1.0 - SUMO Database

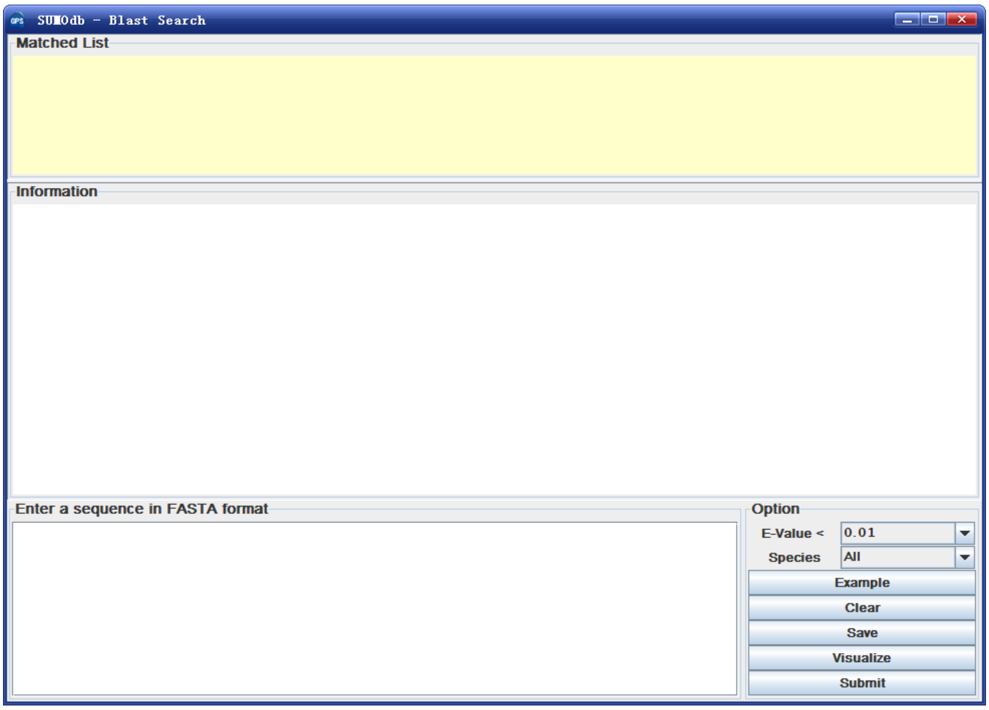
Tag	Content
SUMOdb ID	SUMOdb-3702-00000
Entry Name	F4IL99
UniProt Accession	F4IL99;
Theoretical PI	6.11
Molecular Weight	15992.0
Genbank Protein ID	AEC06518.1;
Genbank Nucleotide ID	CP002685;
Protein Name	Cyclophilin ROC3
Protein Synonyms/Alias	
Gene Name	ROC3
Gene Synonyms/Alias	ROC3;AT2G16600;

Finally, all browsed results could be saved in HTML format.

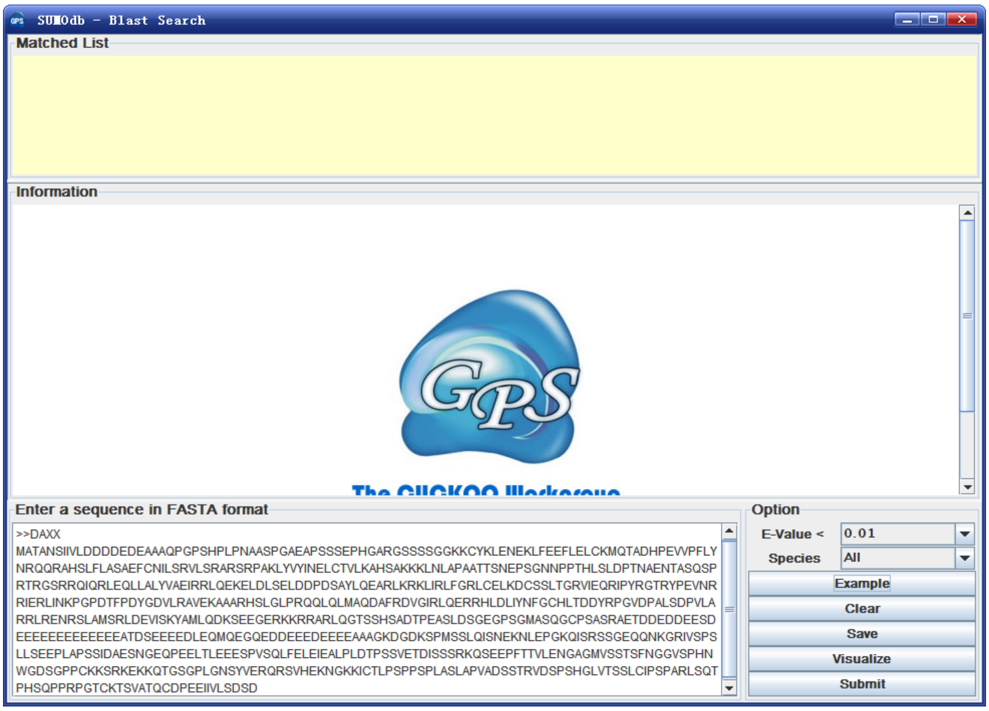
Blast search

The SUMO database also supports the searching function by sequence alignment. The blastp program from NCBI BLAST+ packages was included in SUMO database. Users could input one protein (not mRNA sequence) in FASTA or RAW format a time to search identical or homologous entries. First, users could click on the “Tools” button then click on the “Blast Search” button to open the Blast search window.





Then users could either click on the “Example” button in the Option form or directly input a protein sequence in FASTA or RAW format. Please note that only one protein is permitted a time. Then please click on the “Submit” button to search identical or homologous entries. The E-value and species could be user-defined in the Option form.

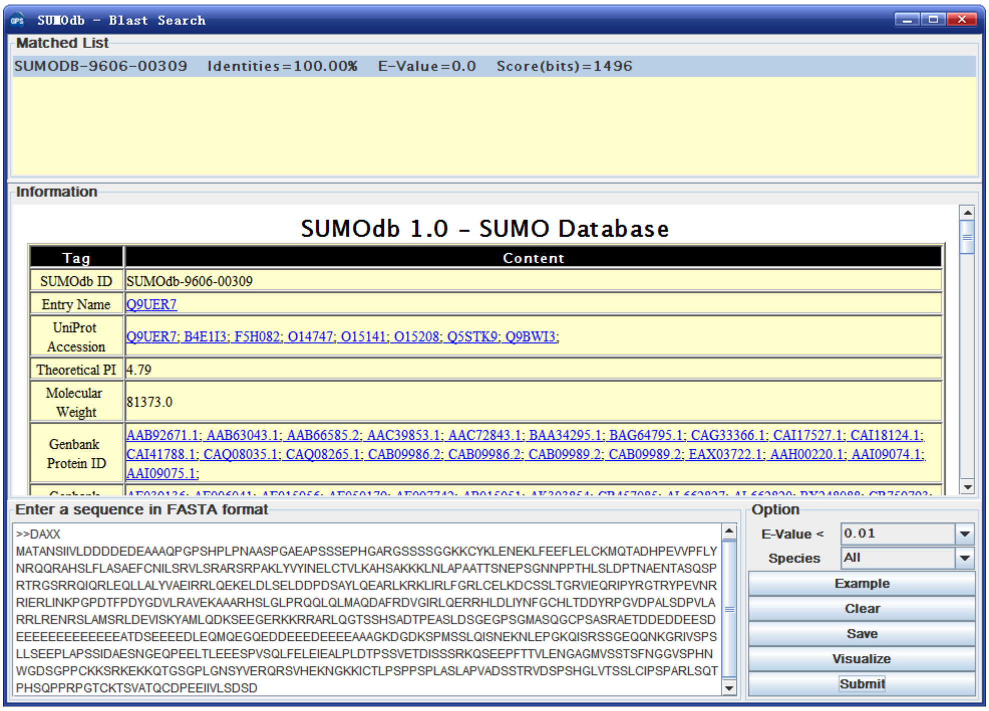


Enter a sequence in FASTA format

```
>>DAXX
MATANSIIVLDDDDEDEAAAQPGPSHPLPNAASPGAEAPSSSEPHGARGSSSSGGKKCYKLENEKLFEEFLELCKMQTADHPEVVPFLY
NRQQRAHSLFLASAEFCNILSRVLSRARSRPAKLYVYINELCTVLKAHSAKKKLNLAPAATTSNEPSGNNPPTHLSLDPTNAENTASQSP
RTRGSRRQIQRLEQLLALYVAEIRRLQEKELDLSELDDPDSAYLQEARLKRKLIRLFGRLCELKDCSSLTGRVIEQRIPYRGTRYPEVNR
RIERLINKPGPDTFPDYGDVLRAVEKAAARHSLGLPRQQLQLMAQDAFRDVGIRLQERRHLDLIYNFGCHLTDDYRPGVDPALSDPVLA
RRLRENRSLAMSRLDEVISKYAMLQDKSEEGERKKRRARLQGTSSHSADTPEASLDSGEGPSGMASQGCPSASRAETDDEDDEESD
EEEEEEEEEEEEEATDSEEEEDLEQMQEGQEDDEEEDEEEEAAAGKDGDKSPMSSLQISNEKNLEPGKQISRSSGEQQNKGRIVSPS
LLSEEPLAPSSIDAESNGEQPEELTLEEESPVSQLFELEIEALPLDTPSSVETDISSSRKQSEEPFTTVLENGAGMVSSTSFNGGVSPHN
WGDSGPPCKKSRKEKKQTGSGPLGNSYVERQRSVHEKNGKKICTLPSPPSPLASLAPVADSSTRVDSPSHGLVTSSLCIPSPARLSQT
PHSQPPRPGTCKTSVATQCDPEEIIVLSDSD
```

E-Value < 0.01

Species All



SUMODB-9606-00309 Identities=100.00% E-Value=0.0 Score(bits)=1496

SUMOdb 1.0 - SUMO Database

Tag	Content
SUMOdb ID	SUMOdb-9606-00309
Entry Name	Q9UER7
UniProt Accession	Q9UER7; B4E1I3; F5H082; O14747; O15141; O15208; Q5STK9; Q9BWI3;
Theoretical PI	4.79
Molecular Weight	81373.0
Genbank Protein ID	AAB92671.1; AAB63043.1; AAB66585.2; AAC39853.1; AAC72843.1; BAA34295.1; BAG64795.1; CAG33366.1; CAI17527.1; CAI18124.1; CAI41788.1; CAQ08035.1; CAQ08265.1; CAB09986.2; CAB09986.2; CAB09989.2; CAB09989.2; EAX03722.1; AAH00220.1; AAI09074.1; AAI09075.1;

Enter a sequence in FASTA format

```
>>DAXX
MATANSIIVLDDDDEDEAAAQPGPSHPLPNAASPGAEAPSSSEPHGARGSSSSGGKKCYKLENEKLFEEFLELCKMQTADHPEVVPFLY
NRQQRAHSLFLASAEFCNILSRVLSRARSRPAKLYVYINELCTVLKAHSAKKKLNLAPAATTSNEPSGNNPPTHLSLDPTNAENTASQSP
RTRGSRRQIQRLEQLLALYVAEIRRLQEKELDLSELDDPDSAYLQEARLKRKLIRLFGRLCELKDCSSLTGRVIEQRIPYRGTRYPEVNR
RIERLINKPGPDTFPDYGDVLRAVEKAAARHSLGLPRQQLQLMAQDAFRDVGIRLQERRHLDLIYNFGCHLTDDYRPGVDPALSDPVLA
RRLRENRSLAMSRLDEVISKYAMLQDKSEEGERKKRRARLQGTSSHSADTPEASLDSGEGPSGMASQGCPSASRAETDDEDDEESD
EEEEEEEEEEEEEATDSEEEEDLEQMQEGQEDDEEEDEEEEAAAGKDGDKSPMSSLQISNEKNLEPGKQISRSSGEQQNKGRIVSPS
LLSEEPLAPSSIDAESNGEQPEELTLEEESPVSQLFELEIEALPLDTPSSVETDISSSRKQSEEPFTTVLENGAGMVSSTSFNGGVSPHN
WGDSGPPCKKSRKEKKQTGSGPLGNSYVERQRSVHEKNGKKICTLPSPPSPLASLAPVADSSTRVDSPSHGLVTSSLCIPSPARLSQT
PHSQPPRPGTCKTSVATQCDPEEIIVLSDSD
```

E-Value < 0.01

Species All

Again, users could visualize any SUMO proteins by clicking on the entries listed in the “Matched list” form. And the results could be saved by clicking on the “Save” button in the Option form. Or you can click on the “Visualize” button to view the SUMO modified sites.

References

1. Geiss-Friedlander, R. and Melchior, F. (2007) Concepts in sumoylation: a decade on. *Nature reviews. Molecular cell biology*, **8**, 947-956.
2. Gill, G. (2005) Something about SUMO inhibits transcription. *Current opinion in genetics & development*, **15**, 536-541.
3. Seeler, J.S. and Dejean, A. (2003) Nuclear and unclear functions of SUMO. *Nature reviews. Molecular cell biology*, **4**, 690-699.
4. Zhao, J. (2007) Sumoylation regulates diverse biological processes. *Cellular and molecular life sciences : CMLS*, **64**, 3017-3033.
5. Raman, N., Nayak, A. and Muller, S. (2013) The SUMO system: a master organizer of nuclear protein assemblies. *Chromosoma*.
6. Lee, L., Sakurai, M., Matsuzaki, S., Arancio, O. and Fraser, P. (2013) SUMO and Alzheimer's Disease. *Neuromolecular medicine*.
7. Eckermann, K. (2013) SUMO and Parkinson's Disease. *Neuromolecular medicine*.
8. Boggio, R. and Chiocca, S. (2006) Viruses and sumoylation: recent highlights. *Current opinion in microbiology*, **9**, 430-436.
9. Wang, J., Chen, L., Wen, S., Zhu, H., Yu, W., Moskowitz, I.P., Shaw, G.M., Finnell, R.H. and Schwartz, R.J. (2011) Defective sumoylation pathway directs congenital heart disease. *Birth defects research. Part A, Clinical and molecular teratology*, **91**, 468-476.
10. Wang, J. and Schwartz, R.J. (2010) Sumoylation and regulation of cardiac gene expression. *Circulation research*, **107**, 19-29.
11. Flotho, A. and Melchior, F. (2013) Sumoylation: A Regulatory Protein Modification in Health and Disease. *Annual review of biochemistry*, **82**.
12. Xue, Y., Ren, J., Gao, X., Jin, C., Wen, L. and Yao, X. (2008) GPS 2.0, a tool to predict kinase-specific phosphorylation sites in hierarchy. *Molecular & cellular proteomics : MCP*, **7**, 1598-1608.

Release Note

1. Jan. 7th, 2008, the online service and the local stand-alone packages of SUMOsp 2.0 were released.
2. Jan. 29th, 2008, a bug was found that the version 2.0 couldn’t be properly used under non-English Operating Systems.
3. Feb. 16th, 2008, the version 2.0 manual was written and included in the packages.
4. Aug. 28th, 2008, DOG (Domain Graph) 1.0 was integrated into SUMOsp 2.0.2 and a new function of visualizing the predicted sites was added.
   Feb. 2nd, 2009，SUMOsp version 2.0.3 was released. We moved the SUMOsp web server to a new website (http://sumosp.biocuckoo.org) and a new GPS logo was put into use.
5. Jul. 23rd, 2009，SUMOsp version 2.0.4 was released. Check for update function was added. DOG (Domain Graph) was updated to version 1.0.5.
6. Dec.30th, 2013, by improving the prediction algorithm with the Particle Swarm Optimization (PSO) and adding the novel SUMO-interaction Motifs prediction feature, we developed an updated version of SUMOsp and renamed it as GPS-SUMO 1.0.